

Improving on the Cut-Set Bound via Geometric Analysis of Typical Sets

Xiugang Wu, Ayfer Özgür and Liang-Liang Xie

Abstract

We consider the discrete memoryless symmetric primitive relay channel, where, a source X wants to send information to a destination Y with the help of a relay Z and the relay can communicate to the destination via an error-free digital link of rate R_0 , while Y and Z are conditionally independent and identically distributed given X . We develop two new upper bounds on the capacity of this channel that are tighter than existing bounds, including the celebrated cut-set bound. Our approach significantly deviates from the standard information-theoretic approach for proving upper bounds on the capacity of multi-user channels. We build on the blowing-up lemma to analyze the probabilistic geometric relations between the typical sets of the n -letter random variables associated with a reliable code for communicating over this channel. These relations translate to new entropy inequalities between the n -letter random variables involved.

As an application of our bounds, we study an open question posed by (Cover, 1987), namely, what is the minimum needed Z - Y link rate R_0^* in order for the capacity of the relay channel to be equal to that of the broadcast cut. We consider the special case when the X - Y and X - Z links are both binary symmetric channels. Our tighter bounds on the capacity of the relay channel immediately translate to tighter lower bounds for R_0^* . More interestingly, we show that when $p \rightarrow 1/2$, $R_0^* \geq 0.1803$; even though the broadcast channel becomes completely noisy as $p \rightarrow 1/2$ and its capacity, and therefore the capacity of the relay channel, goes to zero, a strictly positive rate R_0 is required for the relay channel capacity to be equal to the broadcast bound. Existing upper bounds on the capacity of the relay channel, and the cut-set bound in particular, would rather imply $R_0^* \rightarrow 0$, while achievability schemes require $R_0^* \rightarrow 1$. We conjecture that $R_0^* \rightarrow 1$ as $p \rightarrow 1/2$.

I. INTRODUCTION

Characterizing the capacity of relay channels [3] has been a long-standing open problem in network information theory. The seminal work of Cover and El Gamal [4] has introduced two basic achievability schemes: Decode-and-Forward and Compress-and-Forward, and derived a general upper bound on the capacity of this channel, now known as the cut-set bound. Over the last decade, significant progress has been made on the achievability side: these schemes have been extended and unified to multi-relay networks [5]–[7] and many new relaying strategies have been discovered, such as Amplify-and-Forward, Quantize-Map-and-Forward, Compute-and-Forward, Noisy Network Coding, Hybrid Coding etc [8]–[12]. However, the progress on developing upper bounds that are tighter than the cut-set bound has been relatively limited. In particular, in most of the special cases where the capacity is known, the converse is given by the cut-set bound [4], [13]–[15].

In general, however, the cut-set bound is known to be not tight. Specifically, consider the primitive relay channel depicted in Fig. 1, where the source's input X is received by the relay Z and the destination Y through a channel $p(y, z|x)$, and the relay Z can communicate to the destination Y via an error-free digital link of rate R_0 . When Y and Z are conditionally independent given X , and Y is a stochastically degraded version of Z , Zhang [16] uses the blowing-up lemma [17] to show that the capacity can be

This work was supported in part by the NSF CAREER award 1254786 and by the Center for Science of Information (CSoI), an NSF Science and Technology Center, under grant agreement CCF-0939370. This paper was presented in part at the 2015 IEEE International Symposium on Information Theory [1] and the 2016 International Zurich Seminar on Communications [2].

X. Wu and A. Özgür are with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305, USA (e-mail: x23wu@stanford.edu; aozgur@stanford.edu).

L.-L. Xie is with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1 (e-mail: llxie@uwaterloo.ca).

strictly smaller than the cut-set bound in certain regimes of this channel. However, Zhang's result does not provide any information regarding the gap or suggest a way to compute it. For a special case of the primitive relay channel where the noise for the X - Y link is modulo additive and Z is a corrupted version of this noise, Aleksic, Razaghi and Yu characterize the capacity and show that it is strictly lower than the cut-set bound [18]. While this result provides an exact capacity characterization for a non-trivial special case, it builds strongly on the peculiarity of the channel model and in this respect its scope is more limited than Zhang's result.

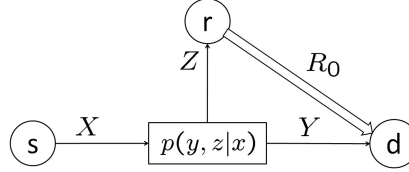


Fig. 1. Primitive relay channel.

More recently, a new upper bound demonstrating an explicit gap to the cut-set bound was developed by Xue [19] for general primitive relay channels. Xue's bound relates the gap of the cut-set bound to the reliability function of the X - Y link. In particular, it builds on the blowing-up lemma to lower bound the successful decoding probability based on Y only and then compare it to the reliability function of the single-user channel X - Y . Unlike Zhang's result, Xue's bound identifies an explicit gap to the cut-set bound that can be numerically computed. However, it also has some obvious drawbacks over the cut-set bound. For example, it bounds only the information flow from the source and the relay to the destination (the multiple-access cut) and ignores the flow from the source to the relay and the destination (the broadcast cut). As such, it can be also looser than the cut-set bound since it does not capture the inherent trade-off between maximizing the information flows across these two different cuts of the network.

In this paper, we present two new upper bounds on the capacity of the primitive relay channel that are generally tighter than the cut-set bound. To simplify exposition, we concentrate on the symmetric case (Y and Z are conditionally independent and identically distributed given X) in this paper, however our results can be extended to the asymmetric case via channel simulation arguments [31]. Just like Zhang and Xue, we critically build on the blowing up lemma, however we develop novel ways for utilizing it which lead to simpler arguments and tighter results. In general, proving an upper bound on the capacity of a multi-user channel involves dealing with entropy relations between the various n -letter random variables induced by the reliable code and the channel structure (together with using Fano's inequality). In order to prove the desired relations between the entropies of the n -letter random variables involved, in this paper we consider their B -length i.i.d. extensions (leading to length B i.i.d. sequences of n -letter random variables). We then use the blowing-up lemma to analyze the geometry of the typical sets associated with these B -length sequences. The key step in our development is to translate the (probabilistic) geometric relations between these typical sets into new entropy relations between the random variables involved. While both of our bounds are based on this same approach, they use different arguments to translate the geometry of the typical sets to entropy relations, and eventually lead to two different bounds on the capacity of the channel which do not include each other in general.

As an application of our bounds, we consider the binary symmetric channel, i.e., we assume both the X - Y and X - Z links are binary symmetric channels with crossover probability, say, p . We demonstrate that both our bounds perform strictly better than the cut-set bound and Xue's bound, and particularly, our second bound provides considerable gain over these earlier bounds. We then use our bounds to investigate an open question posed by Cover [20] which asks for the minimum required Z - Y link rate R_0^* in order for the capacity of the relay channel to be equal to the capacity of the broadcast cut, i.e. $\max_{p(x)} I(X; Y, Z)$. Obviously as R_0^* becomes larger the capacity of the relay channel does approach the capacity of the broadcast cut. For example, in the binary symmetric case if $R_0 = 1$, the relay can convey its noisy observation as it is to the destination, therefore the broadcast cut capacity is trivially achievable. In this

sense, Cover's open problem asks how smaller R_0 can be made than 1 without decreasing the capacity of the relay channel. Interestingly, there is a striking dichotomy between the currently available upper and lower bounds for R_0^* when $p \rightarrow 1/2$, i.e. when the broadcast channel becomes completely noisy and its capacity, and therefore the capacity of the relay channel, goes to zero. Achievability schemes, Hash-and-Forward in particular, require $R_0 \rightarrow 1$ even though the capacity itself tends to zero. The cut-set bound and Xue's bound, on the other hand, require $R_0 \rightarrow 0$ in order for the capacity to be equal to the broadcast capacity. By strengthening our second bound in this specific case, we show that $R_0^* \geq 0.1803$; indeed a strictly positive rate R_0 is needed in order to achieve the vanishing broadcast capacity. We conjecture that $R_0^* \rightarrow 1$ when $p \rightarrow 1/2$; to achieve the broadcast capacity the relay has no choice but to forward its observation, which is almost pure noise, as it is to the destination.

A. Organization of the Paper

The remainder of the paper is organized as follows. Sections II and III introduces the channel model and reviews the existing upper bounds for primitive relay channels, respectively. Section IV discusses our new upper bounds for symmetric primitive relay channels in detail, followed by a treatment on the binary symmetric channel case in Section V. Sections VI, VII and VIII are then dedicated to the proofs of our bounds. Finally, some concluding remarks are included in Section IX.

II. CHANNEL MODEL

Consider a primitive relay channel as depicted in Fig. 1. The source's input X is received by the relay Z and the destination Y through a channel

$$(\Omega_X, p(y, z|x), \Omega_Y \times \Omega_Z)$$

where Ω_X, Ω_Y and Ω_Z are finite sets denoting the alphabets of the source, the destination and the relay, respectively, and $p(y, z|x)$ is the channel transition probability; the relay Z can communicate to the destination Y via an error-free digital link of rate R_0 .

For this channel, a code of rate R for n channel uses, denoted by

$$(\mathcal{C}_{(n,R)}, f_n(z^n), g_n(y^n, f_n(z^n))), \text{ or simply, } (\mathcal{C}_{(n,R)}, f_n, g_n),$$

consists of the following:

- 1) A codebook at the source X ,

$$\mathcal{C}_{(n,R)} = \{x^n(m) \in \Omega_X^n, m \in \{1, 2, \dots, 2^{nR}\}\};$$

- 2) An encoding function at the relay Z ,

$$f_n : \Omega_Z^n \rightarrow \{1, 2, \dots, 2^{nR_0}\};$$

- 3) A decoding function at the destination Y ,

$$g_n : \Omega_Y^n \times \{1, 2, \dots, 2^{nR_0}\} \rightarrow \{1, 2, \dots, 2^{nR}\}.$$

The average probability of error of the code is defined as

$$P_e^{(n)} = \Pr(g_n(Y^n, f_n(Z^n)) \neq M),$$

where the message M is assumed to be uniformly drawn from the message set $\{1, 2, \dots, 2^{nR}\}$. A rate R is said to be achievable if there exists a sequence of codes

$$\{(\mathcal{C}_{(n,R)}, f_n, g_n)\}_{n=1}^{\infty}$$

such that the average probability of error $P_e^{(n)} \rightarrow 0$ as $n \rightarrow \infty$.

The capacity of the primitive relay channel is the supremum of all achievable rates, denoted by $C(R_0)$. Also, denote by C_{XY}, C_{XZ} and C_{XYZ} the capacities of the channels X - Y , X - Z , and X - YZ , respectively. Obviously, we have $C(0) = C_{XY}$ and $C(\infty) = C_{XYZ}$.

A. Symmetric Primitive Relay Channel

In this paper, we focus on the symmetric case of the primitive relay channel, that is, when Y and Z are conditionally independent and identically distributed given X . Formally, a primitive relay channel is said to be symmetric if

$$1) \ p(y, z|x) = p(y|x)p(z|x),$$

$$2) \ \Omega_Y = \Omega_Z := \Omega, \text{ and } \Pr(Y = \omega|X = x) = \Pr(Z = \omega|X = x) \text{ for any } \omega \in \Omega \text{ and } x \in \Omega_X.$$

In this case, we also use $p(\omega|x)$ to denote the transition probability of both the X - Y and X - Z channels.

III. EXISTING UPPER BOUNDS FOR PRIMITIVE RELAY CHANNELS

For general primitive relay channels, the well-known cut-set bound can be stated as follows.

Proposition 3.1 (Cut-set Bound): For the general primitive relay channel, if a rate R is achievable, then there exists some $p(x)$ such that

$$\begin{cases} R \leq I(X; Y, Z) \end{cases} \quad (1)$$

$$\begin{cases} R \leq I(X; Y) + R_0. \end{cases} \quad (2)$$

Inequalities (1) and (2) are generally known as the broadcast bound and multiple-access bound, since they correspond to the broadcast channel X - YZ and multiple-access channel XZ - Y , respectively.

Note that although the cut-set bound in (1)–(2) is tight for most of the cases where the capacity is determined [4], [13]–[15], it is known to be not tight in general. The first counterexample was given by Zhang in [16], where he considered a class of stochastically degraded primitive relay channels. Using the blowing-up lemma [17], he showed that the capacity of the channel can be strictly smaller than the cut-set bound.

Proposition 3.2 (Zhang [16]): For a primitive relay channel, if Y and Z are conditionally independent given X , and Y is a stochastically degraded version of Z (i.e. there exists some $q(y|z)$ such that $p(y|x) = \sum_z p(z|x)q(y|z)$), then the capacity $C(R_0)$ of the channel satisfies

$$C(R_0) < C_{XY} + R_0 \quad (3)$$

when

$$R_0 > \max_{p(x): I(X; Y) = C_{XY}} I(X; Z) - C_{XY}. \quad (4)$$

In the regime where R_0 satisfies both (4) and the condition

$$C_{XY} + R_0 < \max_{p(x): I(X; Y) = C_{XY}} I(X; Y, Z),$$

the cut-set bound becomes $C_{XY} + R_0$, however the strictness of the inequality in (3) implies that the cut-set bound is loose with some positive gap. Roughly speaking this corresponds to the regime where the cut-set bound is limited by the multiple-access bound but the source-relay channel is not strong enough to enable the relay to trivially decode the transmitted message. However, note that Zhang's result does not provide any information about how large the gap can be.

Recently, a new upper bound demonstrating an explicit gap to the cut-set bound was developed by Xue [19]. This new bound was first established for the symmetric case, and then extended to the general asymmetric case employing channel simulation theory [21]–[22]. The proof uses a generalized version of the blowing-up lemma [19] to characterize the successful decoding probability based only on Y and then compares it with the reliability function for the single-user channel X - Y . Xue's bound specialized for the symmetric case is given as follows.

Proposition 3.3 (Xue's Bound): For the symmetric primitive relay channel, if a rate R is achievable, then there exists some $a \in [0, R_0]$ such that

$$\begin{cases} R \leq C_{XY} + R_0 - a \end{cases} \quad (5)$$

$$\begin{cases} E(R) \leq H(\sqrt{a}) + \sqrt{a} \log |\Omega| \end{cases} \quad (6)$$

where $H(r)$ is the binary entropy function, and $E(R)$ is the reliability function for the X - Y link defined as

$$E(R) := \max_{\rho \in [-1, 0)} (-\rho R + \min_{p(x)} E_0(\rho, p(x))) \quad (7)$$

with

$$E_0(\rho, p(x)) := -\log \left[\sum_y \left(\sum_x p(x) p(y|x)^{\frac{1}{1+\rho}} \right)^{1+\rho} \right].$$

It can be seen that Xue's bound modifies the original multiple-access bound (2) by introducing an additional term “ $-a$ ” in (5), where “ a ” is a non-negative auxiliary variable subject to the constraint (6). As noted in [19], this implies that the capacity of the symmetric primitive relay channel is strictly less than $C_{XY} + R_0$ for any $R_0 > 0$. To see this, consider any rate $R > C_{XY}$. Then it follows from [23] that $E(R) > 0$, which forces a to be strictly positive in light of (6), and thus $R < C_{XY} + R_0$ by (5). Since a here is numerically computable, Xue's bound in fact improves over Zhang's result in the sense that it provides a lower bound to the gap of the cut-set bound.

Nevertheless, Xue's bound also has two obvious drawbacks: i) compared to the cut-set bound, it lacks a constraint on the broadcast cut and therefore decouples the information flow over the broadcast and multiple-access cuts of the channel (note that his result can be always amended by including the bound $R \leq C_{XYZ}$, however with such an amendment this bound would have no coupling with those in (5) and (6), which can be potentially coupled through $p(x)$ as done in the cut-set bound); ii) there is no coupling between (5) and (6) which can also benefit from a coupling through the input distribution $p(x)$. Our bounds presented in the next section overcome these drawbacks and further improve on Xue's bound. They are also structurally different from Xue's bound as they involve only basic information measures and do not involve the reliability function.

IV. NEW UPPER BOUNDS FOR SYMMETRIC PRIMITIVE RELAY CHANNELS

This section presents two new upper bounds on the capacity of symmetric primitive relay channels that are generally tighter than the cut-set bound. Before stating our main theorems, in the following subsection we first explain the relation of our new bounds to the cut-set bound.

A. Improving on the Cut-Set Bound

Let the relay's transmission be denoted by $I_n = f_n(Z^n)$. Let us recall the derivation of the cut-set bound. The first step in deriving (1)–(2) is to use Fano's inequality to conclude that

$$nR \leq I(X^n; Y^n, I_n) + n\epsilon.$$

We can then either proceed as

$$\begin{aligned} nR &\leq I(X^n; Y^n, I_n) + n\epsilon \\ &\leq I(X^n; Y^n, Z^n) + n\epsilon \\ &\leq nI(X; Y, Z) + n\epsilon \end{aligned}$$

to obtain the broadcast bound (1), where the second inequality follows from the data processing inequality and the single letterization in the third line can be either done with a time-sharing or fixed composition code argument¹; or we can proceed as

$$\begin{aligned} nR &\leq I(X^n; Y^n, I_n) + n\epsilon \\ &\leq I(X^n; Y^n) + H(I_n|Y^n) - H(I_n|X^n) + n\epsilon \end{aligned} \quad (8)$$

$$\leq nI(X; Y) + nR_0 + n\epsilon \quad (9)$$

¹Note that the time-sharing or the fixed composition code argument for single letterization is needed to preserve the coupling to the second inequality in (9) via X .

to obtain the multiple-access bound (2), where to obtain the last inequality we upper bound $H(I_n|Y^n)$ by nR_0 and use the fact that $H(I_n|X^n)$ is non-negative.

Instead of simply lower bounding $H(I_n|X^n)$ by 0 in the last step, our bounds presented in the next two subsections are based on letting $H(I_n|X^n) = na_n$ and proving a third inequality that forces a_n to be strictly non-zero. This new inequality is based on capturing the tension between how large $H(I_n|Y^n)$ and how small $H(I_n|X^n)$ can be. Intuitively, it is easy to see this tension. Specifically, suppose $H(I_n|X^n) \approx 0$, then roughly speaking, this implies that given the transmitted codeword X^n , there is no ambiguity about I_n , or equivalently, all the Z^n sequences jointly typical with X^n are mapped to the same I_n . Since Y^n and Z^n are statistically equivalent given X^n (they share the same typical set given X^n) this would further imply that I_n can be determined based on Y^n , and therefore $H(I_n|Y^n) \approx 0$. This would force the rate to be even smaller than $I(X; Y)$.

Equivalently, rewriting (8) and (9) as

$$R \leq nI(X; Y) + I(I_n; X^n) - I(I_n; Y^n) + n\epsilon, \quad (10)$$

our approach can be thought of as fixing the first n -letter mutual information to be $I(I_n; X^n) \leq n(R_0 - a_n)$ and controlling the second n -letter mutual information. In doing so, we only build on the Markov chain structure $I_n \leftrightarrow Z^n \leftrightarrow X^n \leftrightarrow Y^n$ and the fact that Z^n and Y^n are conditionally i.i.d. given X^n . In particular, we do not employ the fact that these random variables are associated with a reliable code. Note that this approach of directly studying the relation between the n -letter information measures involved significantly deviates from the standard approach in network information theory for developing converses, where one usually seeks to single letterize such n -letter expressions.

More precisely, we proceed as follows. We fix $H(I_n|X^n) = na_n$ and leave this term as it is in (8), yielding

$$R \leq I(X; Y) + R_0 - a_n + \epsilon.$$

We then prove the following two upper bounds on $I(I_n; X^n) - I(I_n; Y^n)$ in terms of a_n :

$$I(I_n; X^n) - I(I_n; Y^n) \leq n \left[H \left(\sqrt{\frac{a_n \ln 2}{2}} \right) + \sqrt{\frac{a_n \ln 2}{2}} \log(|\Omega| - 1) - a_n \right], \quad (11)$$

where $H(r)$ is the binary entropy function; and

$$I(I_n; X^n) - I(I_n; Y^n) \leq n\Delta \left(p(x), \sqrt{\frac{a_n \ln 2}{2}} \right), \quad (12)$$

where $\Delta \left(p(x), \sqrt{\frac{a_n \ln 2}{2}} \right)$ is a quantity that depends on the input distribution $p(x)$ and a_n , which we will formally define in Section IV-C. These two bounds are obtained via bounding $H(I_n|Y^n)$ and $H(Y^n|I_n)$ in terms of a_n respectively, and combined with (10) they immediately yield new constraints on R .

The heart of our argument is therefore to prove the two bounds in (11) and (12). To accomplish this, we suggest a new set of proof techniques. In particular, we look at the B -letter i.i.d. extensions of the random variables X^n, Y^n, Z^n and I_n and study the geometric relations between their typical sets by using the generalized blowing-up lemma. While we use this same general approach for obtaining (11) and (12), we build on different arguments in each case, which eventually leads to two different bounds on the capacity of the relay channel that do not include each other in general.

B. Via Bounding $H(I_n|Y^n)$

Our first bound builds on bounding $H(I_n|Y^n)$ and it is given by the following theorem.

Theorem 4.1: For the symmetric primitive relay channel, if a rate R is achievable, then there exists some $p(x)$ and

$$a \in \left[0, \min \left\{ R_0, H(Z|X), \frac{2}{\ln 2} \left(\frac{|\Omega| - 1}{|\Omega|} \right)^2 \right\} \right] \quad (13)$$

such that

$$\begin{cases} R \leq I(X; Y, Z) & (14) \\ R \leq I(X; Y) + R_0 - a & (15) \\ R \leq I(X; Y) + H \left(\sqrt{\frac{a \ln 2}{2}} \right) + \sqrt{\frac{a \ln 2}{2}} \log(|\Omega| - 1) - a. & (16) \end{cases}$$

Clearly our bound in Theorem 4.1 implies the cut-set bound in Proposition 3.1. In fact, it can be checked that our bound is *strictly* tighter than the cut-set bound for any $R_0 > 0$. For this, note that (15) will reduce to (2) only if $a = 0$; however, if $a = 0$ then (16) will constrain R by the rate $I(X; Y)$ which is lower than the cut-set bound.

Our bound is also generally tighter than Xue's bound, and since Xue's bound implies Zhang's result [16], so does our bound. In particular, our bound overcomes the drawbacks of Xue's bound that are observed in Section III, and furthermore tightens the constraint (6) on a to (16). By contrasting (14)–(16) to (5)–(6), we note the following improvements:

- 1) Our bound introduces the missing constraint on the broadcast cut (14) and couples it with (15)–(16) through the input distribution $p(x)$.
- 2) The term C_{XY} in (5) is replaced by $I(X; Y)$ in (15). Since the distribution $p(x)$ in Theorem 4.1 has to be chosen to satisfy all the constraints (14)–(16), it may not necessarily maximize $I(X; Y)$, and thus (15) is in general stricter than (5).
- 3) The constraint (6) on a is replaced by (16). To show that the latter is stricter, rewrite it as

$$R - I(X; Y) \leq H \left(\sqrt{\frac{a \ln 2}{2}} \right) + \sqrt{\frac{a \ln 2}{2}} \log(|\Omega| - 1) - a. \quad (17)$$

Note that (6) is active only if $R > C_{XY}$. In Appendix A we show that in this case the L.H.S. (left-hand-side) of (17) is generally greater than that of (6), while the R.H.S. (right-hand-side) of (17) is obviously less than that of (6) for any $a > 0$. Therefore, the constraint (16) is also stricter than (6).

A simple example demonstrating the above improvements is given in Appendix B. The improvements 1) and 2) come from fixed composition code analysis [24] (or alternatively a time-sharing argument), while the key to improvement 3), which accounts for the structural change from (6) to (16), is a new argument for bounding $H(I_n|Y^n)$ instead of analyzing the successful decoding probability based only on Y as done in [19].

C. Via Bounding $H(Y^n|I_n)$

Before presenting our second upper bound, we first define a parameter that will be used in stating the theorem.

Definition 4.1: Given a fixed channel transition probability $p(\omega|x)$, for any $p(x)$ and $d \geq 0$, $\Delta(p(x), d)$ is defined as

$$\Delta(p(x), d) := \max_{\tilde{p}(\omega|x)} H(\tilde{p}(\omega|x)|p(x)) + D(\tilde{p}(\omega|x)||p(\omega|x)|p(x)) - H(p(\omega|x)|p(x)) \quad (18)$$

$$\text{s.t.} \quad \frac{1}{2} \sum_{(x, \omega)} |p(x)\tilde{p}(\omega|x) - p(x)p(\omega|x)| \leq d. \quad (19)$$

In the above, we adopt the notation in [25]. Specifically, $D(\tilde{p}(\omega|x)||p(\omega|x)|p(x))$ is the conditional relative entropy defined as

$$D(\tilde{p}(\omega|x)||p(\omega|x)|p(x)) := \sum_{(x,\omega)} p(x)\tilde{p}(\omega|x) \log \frac{\tilde{p}(\omega|x)}{p(\omega|x)}, \quad (20)$$

$H(\tilde{p}(\omega|x)|p(x))$ is the conditional entropy defined with respect to the joint distribution $p(x)\tilde{p}(\omega|x)$, i.e.,

$$H(\tilde{p}(\omega|x)|p(x)) := - \sum_{(x,\omega)} p(x)\tilde{p}(\omega|x) \log \tilde{p}(\omega|x), \quad (21)$$

and $H(p(\omega|x)|p(x))$ is the conditional entropy similarly defined with respect to $p(x)p(\omega|x)$.

$\Delta(p(x), d)$ can be interpreted as follows: given a random variable $X \sim p(x)$, assume we want to describe a related random variable Y . We use a code designed for the conditional distribution $p(w|x)$ for Y given X , while Y actually comes from a distribution $\tilde{p}(w|x)$. The distribution $\tilde{p}(w|x)$ cannot be too different than the assumed distribution in the sense that the total variation distance between the two joint distributions $p(x)p(w|x)$ and $p(x)\tilde{p}(w|x)$ is bounded by d . $\Delta(p(x), d)$ captures the maximal inefficiency we would incur for having Y come from a different distribution than the one assumed, i.e. it is the maximal number of extra bits we would use when compared to the case where Y comes from the assumed distribution.

It can be easily seen that $\Delta(p(x), d) \geq 0$ for all $p(x)$ and $d \geq 0$, and $\Delta(p(x), d) = 0$ when $d = 0$. Moreover, for any fixed $p(x)$ and $d > 0$, $\Delta(p(x), d) = \infty$ if and only if there exists some x with $p(x) > 0$, and some ω such that $p(\omega|x) = 0$. Thus, a sufficient condition for $\Delta(p(x), d) < \infty$ for all $p(x)$ and $d > 0$ is that the channel transition matrix is *fully connected*, i.e., $p(\omega|x) > 0, \forall (x, \omega) \in \Omega_X \times \Omega$. In this case, $\Delta(p(x), d) \rightarrow 0$ as $d \rightarrow 0$ for any $p(x)$.

Example 4.1: Suppose $p(\omega|x)$ corresponds to a binary symmetric channel with crossover probability $p < 1/2$. We derive $\Delta(p(x), d)$ according to Definition 4.1 in Appendix C and obtain

$$\Delta(p(x), d) = \min \{d, 1 - p\} \log \frac{1 - p}{p}. \quad (22)$$

Interestingly, in this case $\Delta(p(x), d)$ has a simple expression that is independent of $p(x)$.

We are now ready to state our second new upper bound, which is proved by bounding $H(Y^n|I_n)$.

Theorem 4.2: For the symmetric primitive relay channel, if a rate R is achievable, then there exists some $p(x)$ and $a \in [0, \min \{R_0, H(Z|X)\}]$ such that

$$\begin{cases} R \leq I(X; Y, Z) \end{cases} \quad (23)$$

$$\begin{cases} R \leq I(X; Y) + R_0 - a \end{cases} \quad (24)$$

$$\begin{cases} R \leq I(X; Y) + \Delta \left(p(x), \sqrt{\frac{a \ln 2}{2}} \right). \end{cases} \quad (25)$$

Theorem 4.2 also implies the cut-set bound in Propositions 3.1. In particular, when the channels X - Y and X - Z have a fully connected transition matrix, our new bound is *strictly* tighter than the cut-set bound since $\Delta(p(x), d) \rightarrow 0$ as $d \rightarrow 0$ for any $p(x)$ in this case.

It should be pointed out that the bounds in Theorems 4.1 and 4.2 are proved based on essentially different arguments, and they do not include each other in general. For instance, in the case when X - Y and X - Z are binary erasure channels (i.e. $\Pr(Y = x|x) = 1 - p$, and $\Pr(Y = \text{erasure}|x) = p, \forall x \in \{0, 1\}$), $\Delta(p(x), d) = \infty$ for all $p(x)$ and $d > 0$, and thus our second bound reduces to the cut-set bound, but the first bound is still strictly tighter than the cut-set bound; whereas in the case when X - Y and X - Z are binary symmetric channels, our second bound is significantly tighter than both the cut-set bound and the first bound as we will show in the next section.

V. BINARY SYMMETRIC CHANNEL

As an application of the upper bounds stated in Sections III and IV, we consider the case where the channel is binary symmetric, i.e., both the X - Y and X - Z links are binary symmetric channels with crossover probability p . The various upper bounds can be specialized to this case as follows (see Appendix D for derivations).

- Cut-set bound (Prop. 3.1):

$$C(R_0) \leq \min \{1 + H(p * p) - 2H(p), 1 - H(p) + R_0\},$$

where $p_1 * p_2 := p_1(1 - p_2) + p_2(1 - p_1)$.

- Xue's bound (Prop. 3.3):

$$C(R_0) \leq \max_{a \in [0, R_0]} \min \{1 - H(p) + R_0 - a, E^{-1}(H(\sqrt{a}) + \sqrt{a})\},$$

where $E^{-1}(\cdot)$ is the inverse function of $E(R)$.

- Our first bound (Thm. 4.1):

$$C(R_0) \leq \max_{a \in [0, \min\{R_0, H(p), \frac{1}{2 \ln 2}\}]} \min \left\{ 1 + H(p * p) - 2H(p), 1 - H(p) + R_0 - a, \right. \\ \left. 1 - H(p) + H\left(\sqrt{\frac{a \ln 2}{2}}\right) - a \right\}.$$

- Our second bound (Thm. 4.2):

$$C(R_0) \leq \max_{a \in [0, \min\{R_0, H(p), \frac{2}{\ln 2}(1-p)^2\}]} \min \left\{ 1 + H(p * p) - 2H(p), 1 - H(p) + R_0 - a, \right. \\ \left. 1 - H(p) + \sqrt{\frac{a \ln 2}{2}} \log \frac{1-p}{p} \right\}.$$

Fig. 2 plots the above bounds for $p = 0.2$ and $R_0 \in [0.15, 0.21]$. As can be seen, both of our bounds perform strictly better than the cut-set bound and Xue's bound, where the latter two are quite close to each other. Particularly, our second bound provides considerable gain over the other three bounds.

A. Cover's Open Problem on the Critical R_0

Now suppose we want to achieve the rate C_{XYZ} for the relay channel. What is the minimum rate needed for the relay-destination link? This question was posed by Cover [20] and has been open for decades. Formally, we are interested in the critical value

$$R_0^* = \inf \{R_0 : C(R_0) = C_{XYZ}\}.$$

The upper bounds on the capacity of the primitive relay channel presented in the previous sections can be immediately used to develop lower bounds on R_0^* . Note that since Xue's bound is always dominated by our first bound, in the following we only compare the lower bounds on R_0^* implied by our two bounds with that implied by the cut-set bound.

- Cut-set bound (Prop. 3.1):

$$R_0^* \geq H(p * p) - H(p).$$

- Our first bound (Thm. 4.1):

$$R_0^* \geq \min_{H\left(\sqrt{\frac{a \ln 2}{2}}\right) - a \geq H(p * p) - H(p)} H(p * p) - H(p) + a.$$

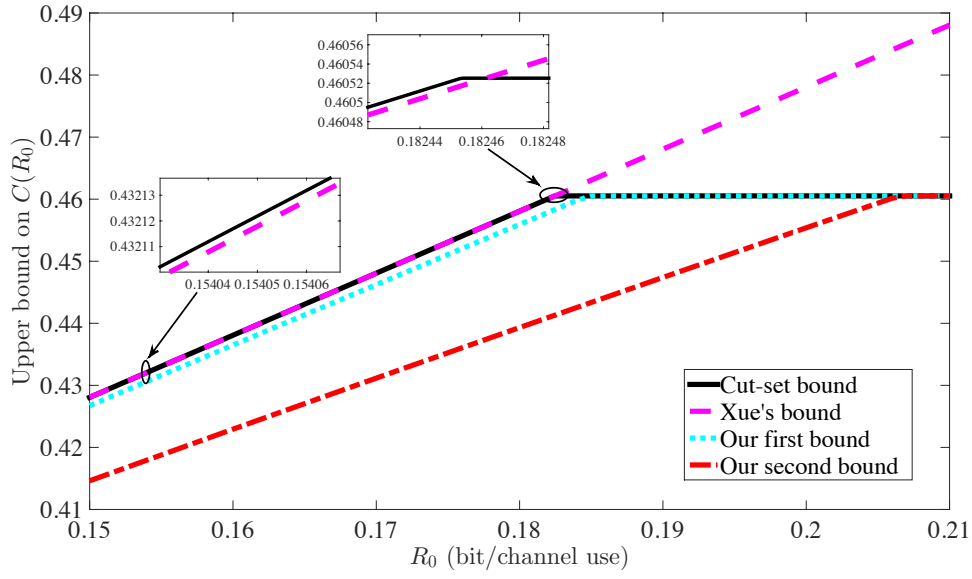


Fig. 2. Upper bounds on $C(R_0)$ for binary symmetric case with $p = 0.2$.

- Our second bound (Thm. 4.2):

$$R_0^* \geq H(p * p) - H(p) + \frac{2}{\ln 2} \left(\frac{H(p * p) - H(p)}{\log \frac{1-p}{p}} \right)^2.$$

Fig. 3 plots these lower bounds on R_0^* versus the crossover probability p . We see again that our second bound provides more gain over the cut-set bound than our first bound does.

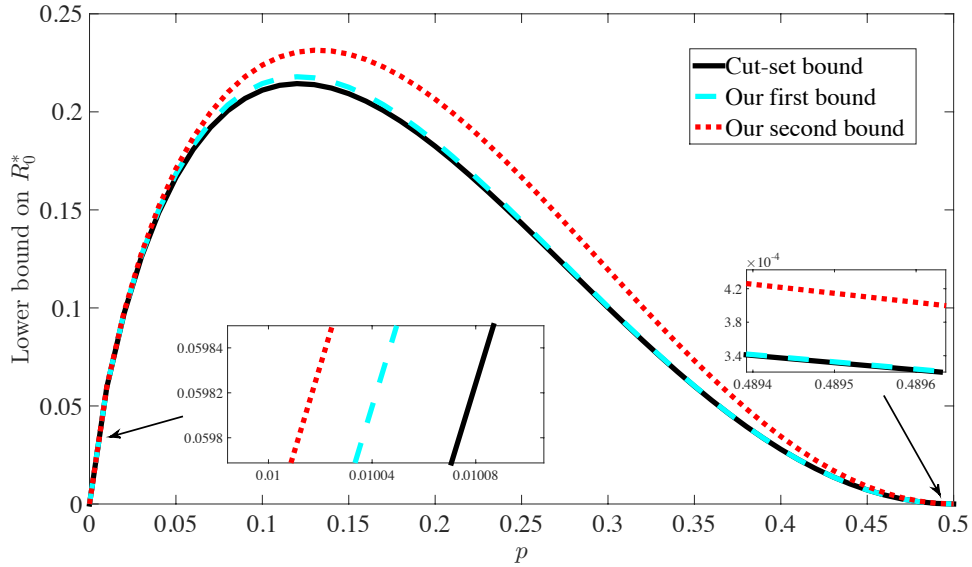


Fig. 3. Lower bounds on R_0^* for binary symmetric case.

From Fig. 3 we observe that all these lower bounds on R_0^* converge to 0 as $p \rightarrow 0$ or $p \rightarrow 1/2$. On the other hand, to achieve C_{XYZ} , a natural way is to use a simple C-F scheme with only Slepian-Wolf binning, a.k.a. Hash-and-Forward (H-F) [26], to faithfully transfer the relay's observation Z^n to the destination so that the joint decoding based on Z^n and Y^n can be performed. This leads to an upper bound on R_0^* , namely $R_0^* \leq H(p * p)$, where $H(p * p)$ is the conditional entropy $H(Z|Y)$ induced by the uniform input distribution. Interestingly, this H-F upper bound also converges to 0 as $p \rightarrow 0$; but as $p \rightarrow 1/2$, it

converges to 1 even though C_{XYZ} is diminishing in this regime, which is in sharp contrast to the above lower bounds on R_0^* that all converge to 0.

This leads to an interesting dichotomy: as $p \rightarrow 1/2$ while achievability requires a full bit of R_0 to support the diminishing C_{XYZ} rate, the converse results allow for a diminishing R_0 . Building on our second upper bound on the capacity of the primitive relay channel, in Section VIII we prove the following improved lower bound on R_0^* , which deviates from 0 as $p \rightarrow 1/2$, thus suggesting that a positive R_0 is needed to achieve C_{XYZ} even when $C_{XYZ} \rightarrow 0$. The proof of this result follows the argument for proving our second bound, however it also critically incorporates the fact that the rate of the codebook is approximately C_{XYZ} in this case as well as the fact that the channel is binary symmetric, which allows us to do a combinatorial geometric analysis of the typical sets in Hamming space.

Theorem 5.1: For the binary symmetric channel case,

$$R_0^* \geq H(p * p) - H(p) + \frac{2}{\ln 2} \left(\frac{H(p * p) - H(p)}{(1 - 2p) \log \frac{1-p}{p}} \right)^2.$$

Proof: See Section VIII. ■

Fig. 4 shows this further improved lower bound on R_0^* as well as the H-F upper bound. Clearly, this improved lower bound is tighter than all other lower bounds, and in particular, it converges to a strictly positive value, 0.1803, as $p \rightarrow 1/2$ while all the other lower bounds converge to 0. This also shows that R_0^* is discontinuous since when $p = 1/2$, the capacity of the relay channel is 0, and therefore trivially $R_0^* = 0$. We indeed believe that $R_0^* \rightarrow 1$ as $p \rightarrow 1/2$ but proving this currently remains out of reach.

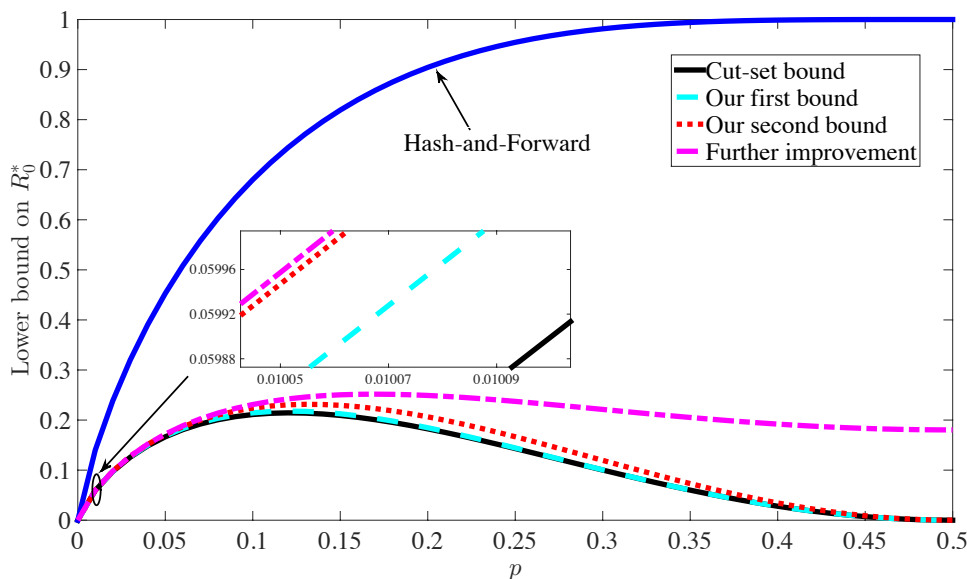


Fig. 4. Improved lower bound on R_0^* for the binary symmetric case.

From Fig. 4, one can observe that in the other extreme, as $p \rightarrow 0$, upper and lower bounds on R_0^* do indeed match and all approach 0. One can indeed check that the speed at which they approach 0 is also not too different. In particular, we can show that H-F is approximately optimal within a multiplicative factor of 2 in the regime where $p \rightarrow 0$. More precisely, letting $R_0^{\text{H-F}} = H(p * p)$ and $R_0^{\text{C-S}} = H(p * p) - H(p)$ denote the H-F bound and the cut-set bound on R_0^* respectively, it can be shown (see Appendix E) that

$$\frac{R_0^{\text{H-F}}}{R_0^{\text{C-S}}} \rightarrow 2 \text{ as } p \rightarrow 0. \quad (26)$$

VI. PROOF OF THEOREM 4.1

In this section, we prove bounds (14)–(16) sequentially with the focus on showing (16).

A. Fixed Composition Code Argument

We start with a fixed composition code argument [24], which is useful for coupling bounds (14)–(16) together through the input distribution $p(x)$. For the purpose of showing Theorem 4.1 such an argument can be replaced by using time sharing random variable [29], however the latter technique is not sufficient in deriving the bound in Theorem 4.2; therefore for consistency, this paper employs the former argument for proving both Theorems 4.1 and 4.2.

Definition 6.1: The composition Q_{x^n} (or empirical probability distribution) of a sequence x^n is the relative proportion of occurrences of each symbol of Ω_X , i.e., $Q_{x^n}(a) = N(a|x^n)/n$ for all $a \in \Omega_X$, where $N(a|x^n)$ is the number of times the symbol a occurs in the sequence x^n .

Definition 6.2: A code for the primitive relay channel is said to be of fixed composition Q , denoted by $(\mathcal{C}_{(n,R)}^{[Q]}, f_n, g_n)$, if all the codewords in $\mathcal{C}_{(n,R)}^{[Q]}$ have the same composition Q .

The following lemma says that if a rate R is achievable by some sequence of codes, then there exists a sequence of fixed composition codes that can achieve essentially the same rate.

Lemma 6.1: Suppose a rate R is achievable over the primitive relay channel. Then for any $\tau > 0$, there exists a sequence of fixed composition codes with rate $R_\tau := R - \tau$

$$\{(\mathcal{C}_{(n,R_\tau)}^{[Q_n]}, f_n, g_n)\}_{n=1}^\infty \quad (27)$$

such that the average probability of error $P_e^{(n)} \rightarrow 0$ as $n \rightarrow \infty$.

Proof: The proof relies on the property that there are only a polynomial number of compositions and can be found in Appendix F. ■

B. Proof of (14)–(15)

To prove Theorem 4.1, in the sequel we will use the reliable fixed composition codes in (27). A benefit of this is that now the various n -letter information quantities have single letter characterizations or bounds, as demonstrated in the following.

Lemma 6.2: For the n -channel use code with fixed composition Q_n , we have

$$\begin{aligned} H(Y^n|X^n) &= H(Z^n|X^n) = nH(Y|X) = nH(Z|X) \\ H(Y^n, Z^n|X^n) &= nH(Y, Z|X), \end{aligned}$$

and

$$\begin{aligned} I(X^n; Y^n) &= I(X^n; Z^n) \leq nI(X; Y) = nI(X; Z) \\ I(X^n; Y^n, Z^n) &\leq nI(X; Y, Z) \end{aligned}$$

where $H(Y|X)$, $H(Z|X)$ and $I(X; Y)$, $I(X; Z)$ are calculated based on $Q_n(x)p(\omega|x)$, and $H(Y, Z|X)$ and $I(X; Y, Z)$ are calculated based on $Q_n(x)p(y|x)p(z|x)$.

Proof: See Appendix G. ■

Let the relay's transmission be denoted by $I_n = f_n(Z^n)$. With the above lemma, we have

$$\begin{aligned} nR_\tau &= n(R - \tau) = H(M) \\ &= I(M; Y^n, I_n) + H(M|Y^n, I_n) \\ &\leq I(X^n; Y^n, I_n) + n\epsilon \\ &\leq I(X^n; Y^n, Z^n) + n\epsilon \\ &\leq n(I(X; Y, Z) + \epsilon) \end{aligned} \quad (28)$$

i.e.,

$$R \leq I(X; Y, Z) + \tau + \epsilon \quad (29)$$

for any $\tau, \epsilon > 0$ and sufficiently large n , where (28) follows from Fano's inequality.

Moreover, for any $\tau, \epsilon > 0$ and sufficiently large n , continuing with (28) we have

$$\begin{aligned}
n(R - \tau) &\leq I(X^n; Y^n, I_n) + n\epsilon \\
&= I(X^n; Y^n) + I(X^n; I_n | Y^n) + n\epsilon \\
&= I(X^n; Y^n) + H(I_n | Y^n) - H(I_n | X^n) + n\epsilon \\
&\leq n(I(X; Y) + R_0 - a_n + \epsilon)
\end{aligned} \tag{30}$$

i.e.,

$$R \leq I(X; Y) + R_0 - a_n + \tau + \epsilon \tag{32}$$

where $a_n := \frac{1}{n}H(I_n | X^n)$ is subject to the following constraint

$$0 \leq a_n \leq \min \left\{ R_0, \frac{1}{n}H(Z^n | X^n) \right\} = \min \{ R_0, H(Z | X) \}. \tag{33}$$

C. Proof of (16)

To prove (16), we continue with (30), and instead of upper bounding $H(I_n | Y^n)$ by nR_0 as in (31), we use the following upper bound on $H(I_n | Y^n)$, whose proof will be given in Section VI-D.

Lemma 6.3: For any fixed n ,

$$H(I_n | Y^n) \leq nV \left(\sqrt{\frac{a_n \ln 2}{2}} \right),$$

with

$$V(r) := \begin{cases} \log |\Omega| & \text{if } r > \frac{|\Omega|-1}{|\Omega|} \\ H(r) + r \log(|\Omega| - 1) & \text{if } r \leq \frac{|\Omega|-1}{|\Omega|}. \end{cases} \tag{34}$$

$$\tag{35}$$

where $H(r)$ is the binary entropy function defined as $H(r) = -r \log r - (1-r) \log(1-r)$.

Plugging the bound on $H(I_n | Y^n)$ in Lemma 6.3 into (30), we have for any $\tau, \epsilon > 0$ and sufficiently large n ,

$$R \leq I(X; Y) + V \left(\sqrt{\frac{a_n \ln 2}{2}} \right) - a_n + \tau + \epsilon \tag{36}$$

Combining (29), (32), (36) and (33), we have that if a rate R is achievable, then for any $\delta > 0$ and sufficiently large n ,

$$\begin{cases} R \leq I(X; Y, Z) + \delta \\ R \leq I(X; Y) + R_0 - a_n + \delta \\ R \leq I(X; Y) + V \left(\sqrt{\frac{a_n \ln 2}{2}} \right) - a_n + \delta \end{cases}$$

where

$$a_n \in [0, \min \{ R_0, H(Z | X) \}].$$

Since δ can be arbitrarily small, we arrive at the following proposition.

Proposition 6.1: If a rate R is achievable, then there exists some $p(x)$ and $a \in [0, \min \{ R_0, H(Z | X) \}]$ such that

$$\begin{cases} R \leq I(X; Y, Z) & (37) \\ R \leq I(X; Y) + R_0 - a & (38) \\ R \leq I(X; Y) + V \left(\sqrt{\frac{a \ln 2}{2}} \right) - a & (39) \end{cases}$$

where

$$V\left(\sqrt{\frac{a \ln 2}{2}}\right) = \begin{cases} \log |\Omega| & \text{if } a > \frac{2}{\ln 2} \left(\frac{|\Omega|-1}{|\Omega|}\right)^2 \\ H\left(\sqrt{\frac{a \ln 2}{2}}\right) + \sqrt{\frac{a \ln 2}{2}} \log(|\Omega| - 1) & \text{if } a \leq \frac{2}{\ln 2} \left(\frac{|\Omega|-1}{|\Omega|}\right)^2. \end{cases}$$

Now we show that Proposition 6.1 is in fact equivalent to Theorem 4.1.

Theorem 4.1 \rightarrow Proposition 6.1: Suppose Theorem 4.1 is true. Then, for any R achievable, there exists some

$$a \in \left[0, \min \left\{ R_0, H(Z|X), \frac{2}{\ln 2} \left(\frac{|\Omega|-1}{|\Omega|}\right)^2 \right\} \right]$$

satisfying (14)–(16). For such $a \leq \frac{2}{\ln 2} \left(\frac{|\Omega|-1}{|\Omega|}\right)^2$, (39) reduces to (16) and thus Proposition 6.1 is also true.

Proposition 6.1 \rightarrow Theorem 4.1: Suppose Proposition 6.1 is true. Then, for any R achievable, there exists some $a \in [0, \min \{R_0, H(Z|X)\}]$ satisfying (37)–(39). If such $a \leq \frac{2}{\ln 2} \left(\frac{|\Omega|-1}{|\Omega|}\right)^2$, then (14)–(16) hold with this a ; otherwise, (14)–(16) hold with the choice of $a' = \frac{2}{\ln 2} \left(\frac{|\Omega|-1}{|\Omega|}\right)^2$. In either case, Theorem 4.1 is also true.

This establishes the equivalence between Proposition 6.1 and Theorem 4.1, and thus completes the proof of Theorem 4.1.

D. Proof of Lemma 6.3

To prove the inequality in Lemma 6.3 for any fixed n , we go to a higher dimensional, say nB dimensional space, to invoke the concepts of typical sets, and resort to a result on measure concentration, namely, the generalized blowing-up lemma.

Specifically, consider the B -length i.i.d. extensions of the random variables X^n, Y^n, Z^n and I_n , i.e.,

$$\{(X^n(b), Y^n(b), Z^n(b), I_n(b))\}_{b=1}^B, \quad (40)$$

where for any $b \in [1 : B]$, $(X^n(b), Y^n(b), Z^n(b), I_n(b))$ has the same distribution as (X^n, Y^n, Z^n, I_n) . For notational convenience, in the sequel we write the B -length vector $[X^n(1), X^n(2), \dots, X^n(B)]$ as \mathbf{X} and similarly define \mathbf{Y}, \mathbf{Z} and \mathbf{I} ; note here we have $\mathbf{I} = [f_n(Z^n(1)), f_n(Z^n(2)), \dots, f_n(Z^n(B))] =: f(\mathbf{Z})$.

The following lemma is critical for establishing Lemma 6.3. We prove this lemma after we finish the proof of Lemma 6.3. The proof is based on typicality arguments combined with the generalized blowing-up lemma.

Lemma 6.4: Let $f^{-1}(\mathbf{i}) := \{\underline{\omega} \in \Omega^{nB} : f(\underline{\omega}) = \mathbf{i}\}$ and $\Gamma_{nB(\sqrt{\frac{a_n \ln 2}{2}} + \delta)}(f^{-1}(\mathbf{i}))$ be its blown-up set defined as

$$\Gamma_{nB(\sqrt{\frac{a_n \ln 2}{2}} + \delta)}(f^{-1}(\mathbf{i})) := \left\{ \underline{\omega} \in \Omega^{nB} : \exists \underline{\omega}' \in f^{-1}(\mathbf{i}) \text{ s.t. } d(\underline{\omega}, \underline{\omega}') \leq nB \left(\sqrt{\frac{a_n \ln 2}{2}} + \delta \right) \right\},$$

where $d(\underline{\omega}, \underline{\omega}')$ denotes the Hamming distance between the two sequences $\underline{\omega}$ and $\underline{\omega}'$. Then for any $\delta > 0$ and B sufficiently large,

$$\Pr(\mathbf{Y} \in \Gamma_{nB(\sqrt{\frac{a_n \ln 2}{2}} + \delta)}(f^{-1}(\mathbf{I}))) \geq 1 - \delta.$$

With the above lemma, we now upper bound $H(\mathbf{I}|\mathbf{Y})$. Let

$$E = \mathbb{I}(\mathbf{Y} \in \Gamma_{nB(\sqrt{\frac{a_n \ln 2}{2}} + \delta)}(f^{-1}(\mathbf{I})))$$

where $\mathbb{I}(\cdot)$ is the indicator function defined as

$$\mathbb{I}(A) = \begin{cases} 1 & \text{if } A \text{ holds} \\ 0 & \text{otherwise.} \end{cases}$$

We have

$$\begin{aligned} H(\mathbf{I}|\mathbf{Y}) &\leq H(\mathbf{I}, E|\mathbf{Y}) \\ &= H(E|\mathbf{Y}) + H(\mathbf{I}|\mathbf{Y}, E) \\ &\leq H(\mathbf{I}|\mathbf{Y}, E) + 1 \\ &= \Pr(E = 1)H(\mathbf{I}|\mathbf{Y}, E = 1) + \Pr(E = 0)H(\mathbf{I}|\mathbf{Y}, E = 0) + 1 \\ &\leq H(\mathbf{I}|\mathbf{Y}, E = 1) + \delta n B R_0 + 1. \end{aligned} \quad (41)$$

To bound $H(\mathbf{I}|\mathbf{Y}, E = 1)$, consider a Hamming ball centered at \mathbf{Y} of radius $nB \left(\sqrt{\frac{a_n \ln 2}{2}} + \delta \right)$, which we denote as²

$$\text{Ball} \left(\mathbf{Y}, nB \left(\sqrt{\frac{a_n \ln 2}{2}} + \delta \right) \right) := \left\{ \underline{\omega} : d(\underline{\omega}, \mathbf{Y}) \leq nB \left(\sqrt{\frac{a_n \ln 2}{2}} + \delta \right) \right\}.$$

The condition $E = 1$, i.e., $\mathbf{Y} \in \Gamma_{nB(\sqrt{\frac{a_n \ln 2}{2}} + \delta)}(f^{-1}(\mathbf{I}))$, ensures that there is at least one point $\underline{\omega} \in f^{-1}(\mathbf{I})$ belonging to this ball, and therefore, given $E = 1$ and \mathbf{Y} the number of different possibilities for \mathbf{I} is bounded by $|\text{Ball}(\mathbf{Y}, nB(\sqrt{\frac{a_n \ln 2}{2}} + \delta))|$, the number of sequences in this Hamming ball, leading to the following upper bound on $H(\mathbf{I}|\mathbf{Y}, E = 1)$,

$$\begin{aligned} H(\mathbf{I}|\mathbf{Y}, E = 1) &\leq \log \left| \text{Ball} \left(\mathbf{Y}, nB \left(\sqrt{\frac{a_n \ln 2}{2}} + \delta \right) \right) \right| \\ &= nB V \left(\sqrt{\frac{a_n \ln 2}{2}} + \delta \right) \end{aligned} \quad (42)$$

$$\leq nB \left[V \left(\sqrt{\frac{a_n \ln 2}{2}} \right) + \delta_1 \right] \quad (43)$$

for some $\delta_1 \rightarrow 0$ as $\delta \rightarrow 0$, where the function $V(\cdot)$ is defined as in (34)–(35), (42) follows from the characterization of the volume of a Hamming ball (see Appendix H for details), and (43) follows from the continuity of the function $V(\cdot)$. Plugging (43) into (41), we have

$$H(\mathbf{I}|\mathbf{Y}) \leq nB \left[V \left(\sqrt{\frac{a_n \ln 2}{2}} \right) + \delta_1 \right] + \delta n B R_0 + 1.$$

Dividing B at both sides of the above inequality and noting that

$$H(\mathbf{I}|\mathbf{Y}) = \sum_{b=1}^B H(I_n(b)|Y^n(b)) = B H(I_n|Y^n),$$

we have

$$H(I_n|Y^n) \leq n \left(V \left(\sqrt{\frac{a_n \ln 2}{2}} \right) + \delta_1 + \delta R_0 + \frac{1}{nB} \right). \quad (44)$$

²The Hamming ball here should be distinguished from the notion of Hamming sphere that will be used later in Section VIII. Specifically, a Hamming ball centered at \mathbf{c} of radius r , denoted by $\text{Ball}(\mathbf{c}, r)$, is defined as the set of points that are within Hamming distance r of \mathbf{c} , whereas a corresponding Hamming sphere, denoted by $\text{Sphere}(\mathbf{c}, r)$, is the set of points that are at a Hamming distance equal to r from \mathbf{c} .

Since δ, δ_1 and $\frac{1}{nB}$ in (44) can all be made arbitrarily small by choosing B sufficiently large, we obtain

$$H(I_n|Y^n) \leq nV \left(\sqrt{\frac{a_n \ln 2}{2}} \right). \quad (45)$$

This finishes the proof of Lemma 6.3.

We are now in a position to prove Lemma 6.4. For this, we will apply the following generalized blowing-up lemma [28, Lemma 12].

Lemma 6.5 (Generalized Blowing-Up Lemma): Let U_1, U_2, \dots, U_n be n independent random variables taking values in a finite set \mathcal{U} . Then, for any $A \subseteq \mathcal{U}^n$ with $\Pr(U^n \in A) \geq 2^{-na_n}$,

$$\Pr(U^n \in \Gamma_{n(\sqrt{\frac{a_n \ln 2}{2}} + r)}(A)) \geq 1 - e^{-2nr^2}, \forall r > 0.$$

Proof of Lemma 6.4: Consider any $(\mathbf{x}, \mathbf{i}) \in \mathcal{T}_\epsilon^{(B)}(X^n, I_n)$, where $\mathcal{T}_\epsilon^{(B)}(X^n, I_n)$ denotes the ϵ -jointly typical sets³ with respect to (X^n, I_n) . From the property of jointly typical sequences (cf. [30, Sec. 2.5]), we have for some $\epsilon_1 \rightarrow 0$ as $\epsilon \rightarrow 0$,

$$p(\mathbf{i}|\mathbf{x}) \geq 2^{-B(H(I_n|X^n) + \epsilon_1)} \geq 2^{-nB(a_n + \epsilon_1)},$$

i.e.,

$$\Pr(\mathbf{Z} \in f^{-1}(\mathbf{i})|\mathbf{x}) \geq 2^{-nB(a_n + \epsilon_1)}.$$

Note that due to the discrete memoryless property of the channel, given \mathbf{x} , \mathbf{Z} is an nB -length sequence of independent random variables, and we then have, by applying Lemma 6.5, that

$$\begin{aligned} \Pr(\mathbf{Z} \in \Gamma_{nB(\sqrt{\frac{a_n \ln 2}{2}} + 2\sqrt{\epsilon_1})}(f^{-1}(\mathbf{i}))|\mathbf{x}) &= \Pr(\mathbf{Z} \in \Gamma_{nB(\sqrt{\frac{(a_n + \epsilon_1) \ln 2}{2}} + [\sqrt{\frac{a_n \ln 2}{2}} + 2\sqrt{\epsilon_1} - \sqrt{\frac{(a_n + \epsilon_1) \ln 2}{2}}])}(f^{-1}(\mathbf{i}))|\mathbf{x}) \\ &\geq \Pr(\mathbf{Z} \in \Gamma_{nB(\sqrt{\frac{(a_n + \epsilon_1) \ln 2}{2}} + [\sqrt{\frac{a_n \ln 2}{2}} + 2\sqrt{\epsilon_1} - \sqrt{\frac{a_n \ln 2}{2}} - \sqrt{\frac{\epsilon_1 \ln 2}{2}}])}(f^{-1}(\mathbf{i}))|\mathbf{x}) \\ &\geq \Pr(\mathbf{Z} \in \Gamma_{nB(\sqrt{\frac{(a_n + \epsilon_1) \ln 2}{2}} + \sqrt{\epsilon_1})}(f^{-1}(\mathbf{i}))|\mathbf{x}) \\ &\geq 1 - e^{-2nB\epsilon_1} \\ &\geq 1 - \sqrt{\epsilon_1} \end{aligned}$$

for sufficiently large B . Noting that \mathbf{Y} and \mathbf{Z} are identically distributed given \mathbf{X} , we obtain

$$\Pr(\mathbf{Y} \in \Gamma_{nB(\sqrt{\frac{a_n \ln 2}{2}} + 2\sqrt{\epsilon_1})}(f^{-1}(\mathbf{i}))|\mathbf{x}) \geq 1 - \sqrt{\epsilon_1},$$

and thus,

$$\begin{aligned} \Pr(\mathbf{Y} \in \Gamma_{nB(\sqrt{\frac{a_n \ln 2}{2}} + 2\sqrt{\epsilon_1})}(f^{-1}(\mathbf{I}))|\mathbf{X}) &= \sum_{(\mathbf{x}, \mathbf{i})} \Pr(\mathbf{Y} \in \Gamma_{nB(\sqrt{\frac{a_n \ln 2}{2}} + 2\sqrt{\epsilon_1})}(f^{-1}(\mathbf{i}))|\mathbf{x}, \mathbf{i}) p(\mathbf{x}, \mathbf{i}) \\ &= \sum_{(\mathbf{x}, \mathbf{i})} \Pr(\mathbf{Y} \in \Gamma_{nB(\sqrt{\frac{a_n \ln 2}{2}} + 2\sqrt{\epsilon_1})}(f^{-1}(\mathbf{i}))|\mathbf{x}) p(\mathbf{x}, \mathbf{i}) \quad (46) \\ &\geq \sum_{(\mathbf{x}, \mathbf{i}) \in \mathcal{T}_\epsilon^{(B)}(X^n, I_n)} \Pr(\mathbf{Y} \in \Gamma_{nB(\sqrt{\frac{a_n \ln 2}{2}} + 2\sqrt{\epsilon_1})}(f^{-1}(\mathbf{i}))|\mathbf{x}) p(\mathbf{x}, \mathbf{i}) \\ &\geq (1 - \sqrt{\epsilon_1}) \sum_{(\mathbf{x}, \mathbf{i}) \in \mathcal{T}_\epsilon^{(B)}(X^n, I_n)} p(\mathbf{x}, \mathbf{i}) \\ &\geq (1 - \sqrt{\epsilon_1})^2 \quad (47) \\ &\geq 1 - 2\sqrt{\epsilon_1} \end{aligned}$$

for sufficiently large B , where (46) follows due to the Markov chain: $\mathbf{Y} \leftrightarrow \mathbf{X} \leftrightarrow \mathbf{Z} \leftrightarrow \mathbf{I}$, and (47) follows since $\Pr(\mathcal{T}_\epsilon^{(B)}(X^n, I_n)) \rightarrow 1$ as $B \rightarrow \infty$. Finally, choosing δ to be $2\sqrt{\epsilon_1}$ concludes the proof of Lemma 6.4. \blacksquare

³This paper adopts the same definitions and notations for typical, jointly typical, and conditionally typical sets as those in [30].

VII. PROOF OF THEOREM 4.2

The bounds (23)–(24) are the same as (14)–(15), which have been proved in Section VI. To show (25), still consider the reliable fixed composition codes in (27). Then we have the following lemma, which upper bounds the conditional entropy $H(Y^n|I_n)$ and whose proof is given in Section VII-A.

Lemma 7.1: For any n -channel use code with fixed composition Q_n ,

$$H(Y^n|I_n) \leq H(X^n|I_n) - H(X^n|Z^n) + nH(Y|X) + n\Delta\left(Q_n, \sqrt{\frac{a_n \ln 2}{2}}\right), \quad (48)$$

where $H(Y|X)$ is calculated based on $Q_n(x)p(\omega|x)$, $a_n = \frac{1}{n}H(I_n|X^n)$, and $\Delta(\cdot, \cdot)$ is as defined in (18)–(19).

With this lemma, we then have

$$\begin{aligned} n(R - \tau) &\leq I(X^n; Y^n, I_n) + n\epsilon \\ &= I(X^n; I_n) + I(X^n; Y^n|I_n) + n\epsilon \\ &= H(X^n) - H(X^n|I_n) + H(Y^n|I_n) - H(Y^n|X^n) + n\epsilon \\ &\leq H(X^n) - H(X^n|I_n) + \left[H(X^n|I_n) - H(X^n|Z^n) + nH(Y|X) + n\Delta\left(Q_n, \sqrt{\frac{a_n \ln 2}{2}}\right) \right] \\ &\quad - H(Y^n|X^n) + n\epsilon \\ &= I(X^n; Z^n) + n\Delta\left(Q_n, \sqrt{\frac{a_n \ln 2}{2}}\right) + n\epsilon \end{aligned} \quad (49)$$

$$\leq n \left[I(X; Y) + \Delta\left(Q_n, \sqrt{\frac{a_n \ln 2}{2}}\right) + \epsilon \right] \quad (50)$$

for any $\tau, \epsilon > 0$ and n sufficiently large, where in (49) we have used the fact that $H(Y^n|X^n) = nH(Y|X)$ (cf. Lemma 6.2), and (50) follows from the symmetry between Y^n and Z^n and Lemma 6.2 again. This proves the bound (25) and hence Theorem 4.2.

A. Proof of Lemma 7.1

The remaining step then is to show the entropy inequality (48) in Lemma 7.1. For this, again we look at the B -length i.i.d. sequence of (X^n, Y^n, Z^n, I_n) , i.e.,

$$(\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \mathbf{I}) := \{(X^n(b), Y^n(b), Z^n(b), I_n(b))\}_{b=1}^B.$$

The following lemma is crucial for proving inequality (48), and its own proof will be given in the next subsection.

Lemma 7.2: For any $\delta > 0$ and B sufficiently large, there exists a set $\mathcal{S}(Y^n, I_n)$ of (\mathbf{y}, \mathbf{i}) pairs such that

$$\Pr((\mathbf{Y}, \mathbf{I}) \in \mathcal{S}(Y^n, I_n)) \geq 1 - \delta,$$

and for any $(\mathbf{y}, \mathbf{i}) \in \mathcal{S}(Y^n, I_n)$,

$$p(\mathbf{y}|\mathbf{i}) \geq 2^{-B(H(X^n|I_n) - H(X^n|Z^n) + nH(Y|X) + n\Delta(Q_n, \sqrt{\frac{a_n \ln 2}{2}}) + \delta)}.$$

We will now use Lemma 7.2 to prove Lemma 7.1. Letting $E = \mathbb{I}((\mathbf{Y}, \mathbf{I}) \in \mathcal{S}(Y^n, I_n))$, we have for any $\delta > 0$ and B sufficiently large,

$$\begin{aligned}
H(\mathbf{Y}|\mathbf{I}) &\leq H(\mathbf{Y}, E|\mathbf{I}) \\
&= H(E|\mathbf{I}) + H(\mathbf{Y}|\mathbf{I}, E) \\
&\leq H(\mathbf{Y}|\mathbf{I}, E) + 1 \\
&= \Pr(E = 1)H(\mathbf{Y}|\mathbf{I}, E = 1) + \Pr(E = 0)H(\mathbf{Y}|\mathbf{I}, E = 0) + 1 \\
&\leq H(\mathbf{Y}|\mathbf{I}, E = 1) + \delta n B \log |\Omega| + 1 \\
&= - \sum_{(\mathbf{y}, \mathbf{i}) \in \mathcal{S}(Y^n, I_n)} p(\mathbf{y}, \mathbf{i}|E = 1) \log p(\mathbf{y}|\mathbf{i}, E = 1) + \delta n B \log |\Omega| + 1 \\
&\leq - \sum_{(\mathbf{y}, \mathbf{i}) \in \mathcal{S}(Y^n, I_n)} p(\mathbf{y}, \mathbf{i}|E = 1) \log p(\mathbf{y}|\mathbf{i}) + \delta n B \log |\Omega| + 1 \\
&\leq B \left[H(X^n|I_n) - H(X^n|Z^n) + nH(Y|X) + n\Delta \left(Q_n, \sqrt{\frac{a_n \ln 2}{2}} \right) + \delta \right] + \delta n B \log |\Omega| + 1.
\end{aligned} \tag{51}$$

Dividing B at both sides of (51) and noticing that $H(\mathbf{Y}|\mathbf{I}) = BH(Y^n|I_n)$, we have

$$H(Y^n|I_n) \leq H(X^n|I_n) - H(X^n|Z^n) + nH(Y|X) + n\Delta \left(Q_n, \sqrt{\frac{a_n \ln 2}{2}} \right) + \delta + \delta n \log |\Omega| + \frac{1}{B}.$$

Since both δ and $\frac{1}{B}$ in the above inequality can be made arbitrarily small by choosing B sufficiently large, Lemma 7.1 is thus proved.

B. Proof of Lemma 7.2

Let $\mathcal{S}(Y^n, I_n)$ be defined as

$$\mathcal{S}(Y^n, I_n) := \{(\mathbf{y}, \mathbf{i}) : \mathbf{y} \in \Gamma_{nB(\sqrt{\frac{a_n \ln 2}{2}} + \epsilon)}(\mathcal{T}_\epsilon^{(B)}(Z^n|\mathbf{i}))\}. \tag{52}$$

We first show that for any $\epsilon > 0$ and B sufficiently large,

$$\Pr((\mathbf{Y}, \mathbf{I}) \in \mathcal{S}(Y^n, I_n)) \geq 1 - \epsilon. \tag{53}$$

For this, consider any $(\mathbf{x}, \mathbf{i}) \in \mathcal{T}_\epsilon^{(B)}(X^n, I_n)$, $\tilde{\epsilon} > 0$. By the joint typicality lemma (cf. [30, Sec. 2.5.1]), we have

$$\begin{aligned}
\Pr(\mathbf{Z} \in \mathcal{T}_\epsilon^{(B)}(Z^n|\mathbf{x}, \mathbf{i})|\mathbf{x}) &\geq 2^{-B(I(Z^n; I_n|X^n) + \tilde{\epsilon}_1)} \\
&= 2^{-B(H(I_n|X^n) + \tilde{\epsilon}_1)} \\
&\geq 2^{-nB(a_n + \tilde{\epsilon}_1)},
\end{aligned}$$

where $\tilde{\epsilon}_1 \rightarrow 0$ as $\tilde{\epsilon} \rightarrow 0$ and $B \rightarrow \infty$. Since $\mathcal{T}_\epsilon^{(B)}(Z^n|\mathbf{x}, \mathbf{i}) \subseteq \mathcal{T}_\epsilon^{(B)}(Z^n|\mathbf{i})$, we further have

$$\Pr(\mathbf{Z} \in \mathcal{T}_\epsilon^{(B)}(Z^n|\mathbf{i})|\mathbf{x}) \geq 2^{-nB(a_n + \tilde{\epsilon}_1)}.$$

Then, by applying Lemma 6.5 along the same lines as the proof of Lemma 6.4, we can obtain

$$\Pr(\mathbf{Y} \in \Gamma_{nB(\sqrt{\frac{a_n \ln 2}{2}} + 2\sqrt{\tilde{\epsilon}_1})}(\mathcal{T}_\epsilon^{(B)}(Z^n|\mathbf{i}))) \geq 1 - 2\sqrt{\tilde{\epsilon}_1}$$

for sufficiently large B . Choosing ϵ to be $\max\{2\sqrt{\tilde{\epsilon}_1}, \tilde{\epsilon}\}$ then proves (53).

Consider any $(\mathbf{y}, \mathbf{i}) \in \mathcal{S}(Y^n, I_n)$. By the definition of $\mathcal{S}(Y^n, I_n)$, we can find one $\mathbf{z} \in \mathcal{T}_\epsilon^{(B)}(Z^n|\mathbf{i})$ such that

$$d(\mathbf{y}, \mathbf{z}) \leq nB \left(\sqrt{\frac{a_n \ln 2}{2}} + \epsilon \right). \quad (54)$$

Then,

$$\begin{aligned} p(\mathbf{y}|\mathbf{i}) &= \sum_{\mathbf{x}} p(\mathbf{y}|\mathbf{x})p(\mathbf{x}|\mathbf{i}) \\ &\geq \sum_{\mathbf{x} \in \mathcal{T}_\epsilon^{(B)}(X^n|\mathbf{z}, \mathbf{i})} p(\mathbf{y}|\mathbf{x})p(\mathbf{x}|\mathbf{i}) \end{aligned} \quad (55)$$

$$\geq 2^{-B(H(X^n|I_n) + \epsilon_1)} \sum_{\mathbf{x} \in \mathcal{T}_\epsilon^{(B)}(X^n|\mathbf{z}, \mathbf{i})} p(\mathbf{y}|\mathbf{x}) \quad (56)$$

$$\geq 2^{-B(H(X^n|I_n) + \epsilon_1)} |\mathcal{T}_\epsilon^{(B)}(X^n|\mathbf{z}, \mathbf{i})| \min_{\mathbf{x} \in \mathcal{T}_\epsilon^{(B)}(X^n|\mathbf{z}, \mathbf{i})} p(\mathbf{y}|\mathbf{x}) \quad (57)$$

$$\geq 2^{-B(H(X^n|I_n) + \epsilon_1)} 2^{B(H(X^n|Z^n) - \epsilon_2)} \min_{\mathbf{x} \in \mathcal{T}_\epsilon^{(B)}(X^n|\mathbf{z}, \mathbf{i})} p(\mathbf{y}|\mathbf{x}), \quad (58)$$

for some $\epsilon_1, \epsilon_2 \rightarrow 0$ as $\epsilon \rightarrow 0$ and $B \rightarrow \infty$, where the \mathbf{z} throughout (55)–(58) is the one belonging to $\mathcal{T}_\epsilon^{(B)}(Z^n|\mathbf{i})$ and satisfying (54), and (56) and (58) follow from the properties of jointly typical sequences (cf. [30, Sec. 2.5]).

We now lower bound $p(\mathbf{y}|\mathbf{x})$ for any $\mathbf{x} \in \mathcal{T}_\epsilon^{(B)}(X^n|\mathbf{z}, \mathbf{i})$. Since $\mathbf{x} \in \mathcal{T}_\epsilon^{(B)}(X^n|\mathbf{z}, \mathbf{i})$, we have $(\mathbf{x}, \mathbf{z}) \in \mathcal{T}_\epsilon^{(B)}(X^n, Z^n)$, i.e., (\mathbf{x}, \mathbf{z}) are jointly typical with respect to the n -letter random variables (X^n, Z^n) . Due to the fixed composition code assumption and the discrete memoryless property of the channel, this can be shown (see Appendix I) to further imply that (\mathbf{x}, \mathbf{z}) are also jointly typical with respect to the single-letter random variables (X, Z) , i.e.,

$$|P_{(\mathbf{x}, \mathbf{z})}(x, \omega) - Q_n(x)p(\omega|x)| \leq \epsilon_3 Q_n(x)p(\omega|x), \quad (59)$$

for some $\epsilon_3 \rightarrow 0$ as $\epsilon \rightarrow 0$, where $P_{(\mathbf{x}, \mathbf{z})}(x, \omega)$ denotes the joint empirical distribution of (\mathbf{x}, \mathbf{z}) with respect to (X, Z) , defined as

$$P_{(\mathbf{x}, \mathbf{z})}(x, \omega) = \frac{1}{nB} N(x, \omega|\mathbf{x}, \mathbf{z})$$

where $N(x, \omega|\mathbf{x}, \mathbf{z})$ denotes the number of times the symbols (x, ω) occur in the sequences (\mathbf{x}, \mathbf{z}) . On the other hand, we show in Appendix J that the bound (54) on the Hamming distance between \mathbf{y} and \mathbf{z} can translate to a bound on the total variation distance between the two empirical distributions $P_{(\mathbf{x}, \mathbf{y})}(x, \omega)$ and $P_{(\mathbf{x}, \mathbf{z})}(x, \omega)$ for any \mathbf{x} , namely,

$$\sum_{(x, \omega)} |P_{(\mathbf{x}, \mathbf{y})}(x, \omega) - P_{(\mathbf{x}, \mathbf{z})}(x, \omega)| \leq \frac{2}{nB} d(\mathbf{y}, \mathbf{z}) \leq 2 \left(\sqrt{\frac{a_n \ln 2}{2}} + \epsilon \right). \quad (60)$$

Combining (59) and (60), we have for some $\epsilon_4 \rightarrow 0$ as $\epsilon \rightarrow 0$,

$$\sum_{(x, \omega)} |P_{(\mathbf{x}, \mathbf{y})}(x, \omega) - Q_n(x)p(\omega|x)| \leq 2 \sqrt{\frac{a_n \ln 2}{2}} + \epsilon_4, \quad (61)$$

or equivalently expressed as

$$\frac{1}{2} \sum_{(x, \omega)} |Q_n(x)P_{\mathbf{y}|\mathbf{x}}(\omega|x) - Q_n(x)p(\omega|x)| \leq \sqrt{\frac{a_n \ln 2}{2}} + \frac{\epsilon_4}{2}, \quad (62)$$

where we have used the fact that the empirical distribution $P_{\mathbf{x}}(x) = Q_n(x)$ due to the fixed composition code assumption, and $P_{\mathbf{y}|\mathbf{x}}(\omega|x)$ is the conditional empirical distribution satisfying

$$N(x, \omega|\mathbf{x}, \mathbf{y}) = N(x|\mathbf{x})P_{\mathbf{y}|\mathbf{x}}(\omega|x).$$

To bound $p(\mathbf{y}|\mathbf{x})$, we have

$$\begin{aligned} -\frac{1}{nB} \log p(\mathbf{y}|\mathbf{x}) &= -\frac{1}{nB} \sum_{i=1}^{nB} \log p(y_i|x_i) \\ &= -\sum_{(x,\omega)} P_{(\mathbf{x},\mathbf{y})}(x, \omega) \log p(\omega|x) \\ &= \sum_{(x,\omega)} [-P_{(\mathbf{x},\mathbf{y})}(x, \omega) \log p(\omega|x) + P_{(\mathbf{x},\mathbf{y})}(x, \omega) \log P_{\mathbf{y}|\mathbf{x}}(\omega|x) - P_{(\mathbf{x},\mathbf{y})}(x, \omega) \log P_{\mathbf{y}|\mathbf{x}}(\omega|x)] \\ &= -\sum_{(x,\omega)} P_{(\mathbf{x},\mathbf{y})}(x, \omega) \log P_{\mathbf{y}|\mathbf{x}}(\omega|x) + \sum_{(x,\omega)} P_{(\mathbf{x},\mathbf{y})}(x, \omega) \log \frac{P_{\mathbf{y}|\mathbf{x}}(\omega|x)}{p(\omega|x)} \\ &= H(P_{\mathbf{y}|\mathbf{x}}(\omega|x)|P_{\mathbf{x}}(x)) + D(P_{\mathbf{y}|\mathbf{x}}(\omega|x)||p(\omega|x)|P_{\mathbf{x}}(x)) \\ &= H(P_{\mathbf{y}|\mathbf{x}}(\omega|x)|Q_n(x)) + D(P_{\mathbf{y}|\mathbf{x}}(\omega|x)||p(\omega|x)|Q_n(x)), \end{aligned} \quad (63)$$

where $P_{\mathbf{y}|\mathbf{x}}(\omega|x)$ satisfies the constraint (62). For any $p(x)$ and $d \geq 0$, define $\Delta(p(x), d)$ as follows:

$$\Delta(p(x), d) := \max_{\tilde{p}(\omega|x)} H(\tilde{p}(\omega|x)|p(x)) + D(\tilde{p}(\omega|x)||p(\omega|x)|p(x)) - H(p(\omega|x)|p(x)) \quad (64)$$

$$\text{s.t.} \quad \frac{1}{2} \sum_{(x,\omega)} |p(x)\tilde{p}(\omega|x) - p(x)p(\omega|x)| \leq d \quad (65)$$

Comparing (63) and (62) to (64) and (65), we have

$$\begin{aligned} -\frac{1}{nB} \log p(\mathbf{y}|\mathbf{x}) &\leq \Delta\left(Q_n, \sqrt{\frac{a_n \ln 2}{2}} + \frac{\epsilon_4}{2}\right) + H(Y|X) \\ &\leq \Delta\left(Q_n, \sqrt{\frac{a_n \ln 2}{2}}\right) + H(Y|X) + \epsilon_5 \end{aligned} \quad (66)$$

for some $\epsilon_5 \rightarrow 0$ as $\epsilon \rightarrow 0$, where $H(Y|X)$ is calculated based on $Q_n(x)p(\omega|x)$, and the last inequality follows since $\Delta(p(x), d)$ is continuous in d for $d > 0$. This combined with (58) yields that for any $(\mathbf{y}, \mathbf{i}) \in \mathcal{S}(Y^n, I_n)$,

$$\begin{aligned} p(\mathbf{y}|\mathbf{i}) &\geq 2^{-B(H(X^n|I_n)+\epsilon_1)} 2^{B(H(X^n|Z^n)-\epsilon_2)} 2^{-nB(\Delta(Q_n, \sqrt{\frac{a_n \ln 2}{2}})+H(Y|X)+\epsilon_5)} \\ &\geq 2^{-B(H(X^n|I_n)-H(X^n|Z^n)+nH(Y|X)+n\Delta(Q_n, \sqrt{\frac{a_n \ln 2}{2}})+\epsilon_6)} \end{aligned}$$

for some $\epsilon_6 \rightarrow 0$ as $\epsilon \rightarrow 0$ and $B \rightarrow \infty$. Finally, choosing $\delta = \max\{\epsilon, \epsilon_6\}$, we have

$$\Pr(\mathbf{Y}, \mathbf{I}) \in \mathcal{S}(Y^n, I_n) \geq 1 - \delta,$$

and for any $(\mathbf{y}, \mathbf{i}) \in \mathcal{S}(Y^n, I_n)$,

$$p(\mathbf{y}|\mathbf{i}) \geq 2^{-B(H(X^n|I_n)-H(X^n|Z^n)+nH(Y|X)+n\Delta(Q_n, \sqrt{\frac{a_n \ln 2}{2}})+\delta)},$$

which concludes the proof of Lemma 7.2.

VIII. PROOF OF THEOREM 5.1

The main idea for proving Theorem 5.1 follows that for Theorem 4.2. In order to highlight the difference, we first look at the parameter $\Delta(p(x), d)$ that plays an important role in the bound in Theorem 4.2 more closely. In Section IV-C, we have indicated that $\Delta(p(x), d)$ can be interpreted as the maximal number of extra bits we would need to compress Y given X , when Y comes from a conditional distribution $\tilde{p}(w|x)$ instead of the assumed distribution $p(w|x)$ and the total variation distance between the two joint distributions is bounded by d . An alternative role that emerges for this quantity in the context of the proof of Theorem 4.2 is the following.

Consider a pair (\mathbf{x}, \mathbf{z}) of nB -length sequences that are jointly typical with respect to $p(x)p(\omega|x)$. We have

$$p(\mathbf{z}|\mathbf{x}) \doteq 2^{-nBH(p(\omega|x)|p(x))}. \quad (67)$$

Let \mathbf{y} be a sequence taking values in the same alphabet as \mathbf{z} and bounded in its Hamming distance to \mathbf{z} by nBd . Theorem 4.2 is based on obtaining a lower bound on the conditional probability of the sequence \mathbf{y} given \mathbf{x} under $p(x)p(\omega|x)$. In particular, in (66), we show that

$$p(\mathbf{y}|\mathbf{x}) \geq 2^{-nB[H(p(\omega|x)|p(x)) + \Delta(p(x), d)]}. \quad (68)$$

Comparing (67) and (68), we can see that $\Delta(p(x), d)$ characterizes the maximum possible exponential decrease from $p(\mathbf{z}|\mathbf{x})$ to $p(\mathbf{y}|\mathbf{x})$ where (\mathbf{x}, \mathbf{z}) is jointly typical with respect to $p(x)p(\omega|x)$ and the Hamming distance between \mathbf{y} and \mathbf{z} is bounded by nBd .

For the binary symmetric channel, i.e. when the conditional distribution $p(\omega|x)$ corresponds to a binary symmetric channel with crossover probability $p < 1/2$, we show in Appendix C that we have the following explicit expression

$$\begin{aligned} \Delta(p(x), d) &= \min \left\{ H(p) + d \log \frac{1-p}{p}, -\log p \right\} - H(p) \\ &= \min \{d, 1-p\} \log \frac{1-p}{p}. \end{aligned} \quad (69)$$

We next provide an alternative way to obtain this expression by resorting to the above interpretation of $\Delta(p(x), d)$. Note that when $p(\omega|x)$ corresponds to a binary symmetric channel with crossover probability $p < 1/2$, for a (\mathbf{x}, \mathbf{z}) pair that is jointly typical with respect to $p(x)p(\omega|x)$, we have $d(\mathbf{x}, \mathbf{z}) \leq nB(p + \epsilon)$ and

$$p(\mathbf{z}|\mathbf{x}) \doteq 2^{-nBH(p)}. \quad (70)$$

If \mathbf{y} satisfies $d(\mathbf{y}, \mathbf{z}) \leq nBd$, then by the triangle inequality we have

$$\begin{aligned} d(\mathbf{x}, \mathbf{y}) &\leq d(\mathbf{x}, \mathbf{z}) + d(\mathbf{y}, \mathbf{z}) \\ &\leq nB(p + d + 2\epsilon) \end{aligned}$$

and therefore,

$$\begin{aligned} p(\mathbf{y}|\mathbf{x}) &\geq p^{nB(p+d)}(1-p)^{nB-nB(p+d)} \\ &= 2^{-nB[H(p)+d \log \frac{1-p}{p}]}. \end{aligned}$$

Since we also trivially have $p(\mathbf{y}|\mathbf{x}) \geq p^{nB} = 2^{nB \log p}$, it follows that

$$p(\mathbf{y}|\mathbf{x}) \geq 2^{-nB \min \{H(p)+d \log \frac{1-p}{p}, -\log p\}}. \quad (71)$$

Comparing (70) with (71), we have the maximum possible exponential decrease from $p(\mathbf{z}|\mathbf{x})$ to $p(\mathbf{y}|\mathbf{x})$ given by (69).

The above discussion reveals that the proof of Theorem 4.2 inherently uses the triangle inequality to obtain a worst case bound, equal to $nB(d+p)$, on the distance between \mathbf{x} and \mathbf{y} . The new ingredient in the proof of Theorem 5.1 is a more precise analysis on the distance between \mathbf{y} and \mathbf{x} by building on the fact that the capacity of the primitive relay channel is equal to the broadcast bound in the context of Cover's open problem. Specifically, we show that most of the typical \mathbf{x} 's are within a distance $nB(d * p)$ from \mathbf{y} in this case. The detailed proof of Theorem 5.1 is as follows, where we only emphasize the difference from that of Theorem 4.2.

We start by observing that Theorem 5.1 follows from the following proposition as a corollary.

Proposition 8.1: In the binary symmetric channel case, for any $\tau > 0$, if a rate $R = C_{XYZ} - \tau$ is achievable, then there exists some $a \geq 0$ such that

$$\begin{cases} C_{XYZ} - \tau \leq C_{XY} + R_0 - a \\ C_{XYZ} - \tau \leq C_{XY} + \Delta' \left(\sqrt{\frac{a \ln 2}{2}} \right) + \mu \end{cases} \quad (72)$$

$$\quad (73)$$

where

$$\Delta'(d) := d(1-2p) \log \frac{1-p}{p} \quad (74)$$

and $\mu \rightarrow 0$ as $\tau \rightarrow 0$.

Specifically, we have by (72),

$$\begin{aligned} R_0 &\geq C_{XYZ} - C_{XY} + a - \tau \\ &= 1 + H(p * p) - 2H(p) - (1 - H(p)) + a - \tau \\ &= H(p * p) - H(p) + a - \tau, \end{aligned} \quad (75)$$

where we use the fact that $C_{XYZ} = 1 + H(p * p) - 2H(p)$ and $C_{XY} = 1 - H(p)$, and by (73),

$$\begin{aligned} \Delta' \left(\sqrt{\frac{a \ln 2}{2}} \right) &= \sqrt{\frac{a \ln 2}{2}} (1-2p) \log \frac{1-p}{p} \\ &\geq C_{XYZ} - C_{XY} - \tau - \mu \\ &= H(p * p) - H(p) - \tau - \mu \end{aligned}$$

so that

$$a \geq \frac{2}{\ln 2} \left(\frac{H(p * p) - H(p)}{(1-2p) \log \frac{1-p}{p}} \right)^2 - \mu_1 \quad (76)$$

for some $\mu_1 \rightarrow 0$ as $\tau \rightarrow 0$. Combining (75) and (76), we have

$$R_0 \geq H(p * p) - H(p) + \frac{2}{\ln 2} \left(\frac{H(p * p) - H(p)}{(1-2p) \log \frac{1-p}{p}} \right)^2 - \tau - \mu_1,$$

and by the definition of R_0^* ,

$$\begin{aligned} R_0^* &\geq \lim_{\tau \rightarrow 0} H(p * p) - H(p) + \frac{2}{\ln 2} \left(\frac{H(p * p) - H(p)}{(1-2p) \log \frac{1-p}{p}} \right)^2 - \tau - \mu_1 \\ &= H(p * p) - H(p) + \frac{2}{\ln 2} \left(\frac{H(p * p) - H(p)}{(1-2p) \log \frac{1-p}{p}} \right)^2 \end{aligned}$$

which is Theorem 5.1.

We now show Proposition 8.1, whose proof builds on the technique developed to prove Theorem 4.2 but doesn't require fixed composition code analysis. To show (72), along the lines of the proof of (24), we have for any achievable rate $R = C_{XYZ} - \tau$, $\tau > 0$,

$$\begin{aligned} n(C_{XYZ} - \tau) &\leq I(X^n; Y^n) + nR_0 - na_n + n\epsilon \\ &\leq n(C_{XY} + R_0 - a_n + \epsilon), \end{aligned} \quad (77)$$

i.e.,

$$C_{XYZ} - \tau \leq C_{XY} + R_0 - a_n + \epsilon, \quad (78)$$

where (77) follows from the memoryless property of the channel and $\epsilon \rightarrow 0$ as $n \rightarrow \infty$. To show (73), we need the following lemma, whose proof is given in Section VIII-A.

Lemma 8.1: In the binary symmetric channel case, for any n -channel use code with rate $R = C_{XYZ} - \tau$ and $P_e^{(n)} \rightarrow 0$,

$$H(Y^n|I_n) \leq H(X^n|I_n) - H(X^n|Z^n) + nH(Y|X) + n\Delta' \left(\sqrt{\frac{a_n \ln 2}{2}} \right) + n\mu, \quad (79)$$

where μ can be made arbitrarily small by choosing n sufficiently large and τ sufficiently small, $H(Y|X) = H(p)$, $a_n = \frac{1}{n}H(I_n|X^n)$, and $\Delta'(\cdot)$ is as defined in (74).

With the above lemma, following the lines that lead to (49) from (48), we can show that for any achievable rate $R = C_{XYZ} - \tau$, $\tau > 0$,

$$\begin{aligned} n(C_{XYZ} - \tau) &\leq I(X^n; Y^n) + n\Delta' \left(\sqrt{\frac{a_n \ln 2}{2}} \right) + n\mu + n\epsilon \\ &\leq n \left[C_{XY} + \Delta' \left(\sqrt{\frac{a_n \ln 2}{2}} \right) + \mu + \epsilon \right], \end{aligned}$$

i.e.,

$$C_{XYZ} - \tau \leq C_{XY} + \Delta' \left(\sqrt{\frac{a_n \ln 2}{2}} \right) + \mu + \epsilon, \quad (80)$$

where $\mu \rightarrow 0$ as $n \rightarrow \infty$ and $\tau \rightarrow 0$, and $\epsilon \rightarrow 0$ as $n \rightarrow \infty$. Combining (78) and (80) proves Proposition 8.1.

A. Proof of Lemma 8.1

To show the entropy inequality (79) in Lemma 8.1, we again look at the B -length i.i.d. sequence of the n -letter random variables (X^n, Y^n, Z^n, I_n) that are induced by the n -channel use reliable code of rate $C_{XYZ} - \tau$, denoted by

$$(\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \mathbf{I}) := \{(X^n(b), Y^n(b), Z^n(b), I_n(b))\}_{b=1}^B.$$

Lemma 8.2: Given any $\delta > 0$, for τ sufficiently small and n, B sufficiently large, there exists a set $\mathcal{S}(Y^n, I_n)$ of (\mathbf{y}, \mathbf{i}) pairs such that

$$\Pr((\mathbf{Y}, \mathbf{I}) \in \mathcal{S}(Y^n, I_n)) \geq 1 - \delta,$$

and for any $(\mathbf{y}, \mathbf{i}) \in \mathcal{S}(Y^n, I_n)$,

$$p(\mathbf{y}|\mathbf{i}) \geq 2^{-B(H(X^n|I_n) - H(X^n|Z^n) + nH(Y|X) + n\Delta'(\sqrt{\frac{a_n \ln 2}{2}}) + n\delta)}. \quad (81)$$

With Lemma 8.2, along the same lines as in the proof of Lemma 7.1, we can show that

$$H(Y^n|I_n) \leq H(X^n|I_n) - H(X^n|Z^n) + nH(Y|X) + n\Delta' \left(\sqrt{\frac{a_n \ln 2}{2}} \right) + n\delta + \delta n \log |\Omega| + \frac{1}{B},$$

where δ can be made arbitrarily small by choosing τ sufficiently small and n, B sufficiently large. This proves the entropy inequality (79).

We are now in a position to show Lemma 8.2.

Proof of Lemma 8.2: The only difference of Lemma 8.2 from Lemma 7.2 is that here the lower bound on $p(\mathbf{y}|\mathbf{i}), \forall (\mathbf{y}, \mathbf{i}) \in \mathcal{S}(Y^n, I_n)$ is sharpened to that of (81). In particular, assume $\mathcal{S}(Y^n, I_n)$ is defined exactly as in (52). Then, for any specific $(\mathbf{y}_0, \mathbf{i}_0) \in \mathcal{S}(Y^n, I_n)$, we can find one $\mathbf{z}_0 \in \mathcal{T}_\epsilon^{(B)}(Z^n|\mathbf{i})$ such that

$$d(\mathbf{y}_0, \mathbf{z}_0) \leq nB \left(\sqrt{\frac{a_n \ln 2}{2}} + \epsilon \right). \quad (82)$$

The key to the aforementioned sharpening is a tighter upper bound on the distance between \mathbf{y}_0 and the \mathbf{x} 's typical with \mathbf{z}_0 , as stated in the following lemma. The proof of this lemma is based on a combinatorial geometric argument, and is deferred until we finish the proof of Lemma 8.2.

Lemma 8.3: Consider any \mathbf{y}_0 such that $d(\mathbf{y}_0, \mathbf{z}_0) = nBd_0$ for some $\mathbf{z}_0 \in \mathcal{T}_\epsilon^{(B)}(Z^n)$. There exists some $\epsilon' \rightarrow 0$ as $\epsilon \rightarrow 0$ such that

$$\Pr(d(\mathbf{X}, \mathbf{y}_0) \leq nB(d_0 * p + \epsilon') | \mathbf{z}_0) \geq 1 - v$$

where v can be made arbitrarily small by choosing n, B sufficiently large and τ sufficiently small.

Due to the above lemma, we have for some $\epsilon' \rightarrow 0$ as $\epsilon \rightarrow 0$,

$$\begin{aligned} & \Pr(\mathbf{X} \in \text{Ball}(\mathbf{y}_0, nB((\sqrt{\frac{a_n \ln 2}{2}} + \epsilon) * p + \epsilon')) \cap \mathcal{T}_\epsilon^{(B)}(X^n | \mathbf{z}_0, \mathbf{i}_0) | \mathbf{z}_0) \\ & \geq 1 - \Pr(\mathbf{X} \notin \text{Ball}(\mathbf{y}_0, nB((\sqrt{\frac{a_n \ln 2}{2}} + \epsilon) * p + \epsilon')) | \mathbf{z}_0) - \Pr(\mathbf{X} \notin \mathcal{T}_\epsilon^{(B)}(X^n | \mathbf{z}_0, \mathbf{i}_0) | \mathbf{z}_0) \\ & \geq 1 - v - v \\ & = 1 - 2v, \end{aligned}$$

where we have used the fact that $\Pr(\mathbf{X} \notin \mathcal{T}_\epsilon^{(B)}(X^n | \mathbf{z}_0, \mathbf{i}_0) | \mathbf{z}_0) \rightarrow 0$ as $B \rightarrow \infty$. Since for any $\mathbf{x} \in \mathcal{T}_\epsilon^{(B)}(X^n | \mathbf{z}_0, \mathbf{i}_0)$, $p(\mathbf{x} | \mathbf{z}_0) \leq 2^{-B(H(X^n | Z^n) - \epsilon_1)}$ for some $\epsilon_1 \rightarrow 0$ as $\epsilon \rightarrow 0$, we have

$$|\text{Ball}(\mathbf{y}_0, nB((\sqrt{\frac{a_n \ln 2}{2}} + \epsilon) * p + \epsilon')) \cap \mathcal{T}_\epsilon^{(B)}(X^n | \mathbf{z}_0, \mathbf{i}_0)| \geq (1 - 2v) 2^{B(H(X^n | Z^n) - \epsilon_1)} \geq 2^{B(H(X^n | Z^n) - \epsilon_1 - v_1)},$$

where $v_1 \rightarrow 0$ as $\tau \rightarrow 0$ and $n, B \rightarrow \infty$. Therefore,

$$\begin{aligned} p(\mathbf{y}_0 | \mathbf{i}_0) & \geq \sum_{\mathbf{x} \in \text{Ball}(\mathbf{y}_0, nB((\sqrt{\frac{a_n \ln 2}{2}} + \epsilon) * p + \epsilon')) \cap \mathcal{T}_\epsilon^{(B)}(X^n | \mathbf{z}_0, \mathbf{i}_0)} p(\mathbf{y}_0 | \mathbf{x}) p(\mathbf{x} | \mathbf{i}_0) \\ & \geq 2^{B(H(X^n | Z^n) - \epsilon_1 - v_1)} 2^{-B(H(X^n | I_n) + \epsilon_2)} \min_{\mathbf{x} \in \text{Ball}(\mathbf{y}_0, nB((\sqrt{\frac{a_n \ln 2}{2}} + \epsilon) * p + \epsilon'))} p(\mathbf{y}_0 | \mathbf{x}) \end{aligned} \quad (83)$$

where $\epsilon_2 \rightarrow 0$ as $\epsilon \rightarrow 0$. For any \mathbf{x} with $d(\mathbf{x}, \mathbf{y}_0) \leq nB((\sqrt{\frac{a_n \ln 2}{2}} + \epsilon) * p + \epsilon') =: nB(\sqrt{\frac{a_n \ln 2}{2}} * p + \epsilon_3)$, we have

$$\begin{aligned}
p(\mathbf{y}_0|\mathbf{x}) &= (1-p)^{nB-d(\mathbf{x}, \mathbf{y}_0)} p^{d(\mathbf{x}, \mathbf{y}_0)} \\
&= (1-p)^{nB(1-p)} p^{nBp} \cdot \frac{(1-p)^{nB-d(\mathbf{x}, \mathbf{y}_0)} p^{d(\mathbf{x}, \mathbf{y}_0)}}{(1-p)^{nB(1-p)} p^{nBp}} \\
&= 2^{-nBH(p)} \cdot \left(\frac{p}{1-p}\right)^{d(\mathbf{x}, \mathbf{y}_0)-nBp} \\
&\geq 2^{-nBH(p)} \cdot \left(\frac{p}{1-p}\right)^{nB(\sqrt{\frac{a_n \ln 2}{2}} * p + \epsilon_3) - nBp} \\
&= 2^{-nBH(p)} \cdot \left(\frac{p}{1-p}\right)^{nB(\sqrt{\frac{a_n \ln 2}{2}}(1-2p) + \epsilon_3)} \\
&= 2^{-nB(H(p) + (\sqrt{\frac{a_n \ln 2}{2}}(1-2p) + \epsilon_3) \log \frac{1-p}{p})} \\
&= 2^{-nB(H(p) + \Delta'(\sqrt{\frac{a_n \ln 2}{2}}) + \epsilon_4)}
\end{aligned} \tag{84}$$

where $\epsilon_3, \epsilon_4 \rightarrow 0$ as $\epsilon \rightarrow 0$. Plugging (84) into (83), we obtain that

$$p(\mathbf{y}_0|\mathbf{i}_0) \geq 2^{-B(H(X^n|I_n) - H(X^n|Z^n) + nH(p) + n\Delta'(\sqrt{\frac{a_n \ln 2}{2}}) + v_1 + \epsilon_1 + \epsilon_2 + n\epsilon_4)},$$

which proves the lemma. ■

B. Proof of Lemma 8.3

Consider a specific $(\mathbf{y}_0, \mathbf{z}_0)$ pair where $\mathbf{z}_0 \in \mathcal{T}_\epsilon^{(B)}(Z^n)$ and $d(\mathbf{y}_0, \mathbf{z}_0) = nBd_0$. Let

$$q_d := \Pr(d(\mathbf{X}, \mathbf{y}_0) \geq nBd|\mathbf{z}_0).$$

To show Lemma 8.3, we show that there exists some $\epsilon' \rightarrow 0$ as $\epsilon \rightarrow 0$ such that for $d = d_0 * p + \epsilon'$, $q_d \leq v$, for some v satisfying $\lim_{\tau \rightarrow 0} \lim_{n \rightarrow \infty} \lim_{B \rightarrow \infty} v = 0$.

First, using the properties of jointly typical sequences, we can show (see Appendix K) that there exists some $\epsilon_0 \rightarrow 0$ as $\epsilon \rightarrow 0$ such that for any $\delta > 0$ and B sufficiently large,

$$\Pr(p(\mathbf{Y}|\mathbf{z}_0) \leq 2^{-B(H(Y^n|Z^n) - \epsilon_0)}, d(\mathbf{Y}, \mathbf{z}_0) \in [nB(p * p - \epsilon_0), nB(p * p + \epsilon_0)]|\mathbf{z}_0) \geq 1 - \delta. \tag{85}$$

Then consider the following inequalities:

$$\begin{aligned}
\Pr(d(\mathbf{Y}, \mathbf{y}_0) \geq nB(d * p - \epsilon_0)|\mathbf{z}_0) &= \sum_{\mathbf{x}} \Pr(d(\mathbf{Y}, \mathbf{y}_0) \geq nB(d * p - \epsilon_0)|\mathbf{x}) p(\mathbf{x}|\mathbf{z}_0) \\
&\geq \sum_{\mathbf{x}: d(\mathbf{x}, \mathbf{y}_0) \geq nBd} \Pr(d(\mathbf{Y}, \mathbf{y}_0) \geq nB(d * p - \epsilon_0)|\mathbf{x}) p(\mathbf{x}|\mathbf{z}_0) \\
&\geq q_d \cdot \min_{\mathbf{x}: d(\mathbf{x}, \mathbf{y}_0) \geq nBd} \Pr(d(\mathbf{Y}, \mathbf{y}_0) \geq nB(d * p - \epsilon_0)|\mathbf{x}).
\end{aligned} \tag{86}$$

Without loss of generality, consider a specific pair $(\mathbf{x}, \mathbf{y}_0)$ as shown in Fig. 5, where $\mathbf{y}_0 = \mathbf{0}$ and $d(\mathbf{x}, \mathbf{y}_0) = nBd_1 \geq nBd$. By the law of large numbers, we have for any $\delta > 0$ and B sufficiently large,

$$\begin{aligned}
1 - \delta &\leq \Pr\left(\frac{1}{nB} N(0, 1|\mathbf{x}, \mathbf{Y}) \geq \frac{1}{nB} N(0|\mathbf{x})p - \frac{\epsilon_0}{2}, \frac{1}{nB} N(1, 1|\mathbf{x}, \mathbf{Y}) \geq \frac{1}{nB} N(1|\mathbf{x})(1-p) - \frac{\epsilon_0}{2}|\mathbf{x}\right) \\
&\leq \Pr(N(1|\mathbf{Y}) \geq nB(d_1 * p - \epsilon_0)|\mathbf{x}) \\
&\leq \Pr(N(1|\mathbf{Y}) \geq nB(d * p - \epsilon_0)|\mathbf{x}) \\
&= \Pr(d(\mathbf{Y}, \mathbf{y}_0) \geq nB(d * p - \epsilon_0)|\mathbf{x}),
\end{aligned} \tag{87}$$

where (87) follows since the event $N(1|\mathbf{Y}) \geq nB(d_1 * p - \epsilon_0)$ implies $N(1|\mathbf{Y}) \geq nB(d * p - \epsilon_0)$ due to the relation $d_1 \geq d$. Plugging this into (86), we obtain

$$\Pr(d(\mathbf{Y}, \mathbf{y}_0) \geq nB(d * p - \epsilon_0) | \mathbf{z}_0) \geq q_d(1 - \delta). \quad (88)$$

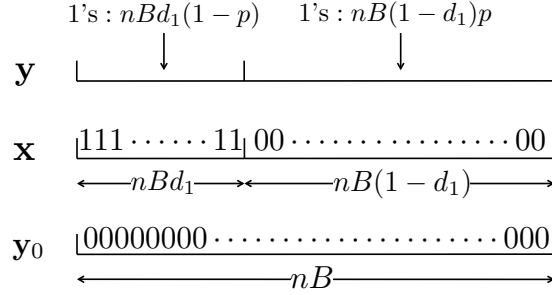


Fig. 5. Illustration of a specific pair $(\mathbf{x}, \mathbf{y}_0)$.

Combining (85) and (88), we have for any $\delta > 0$ and B sufficiently large,

$$\begin{aligned} & \Pr(d(\mathbf{Y}, \mathbf{y}_0) \geq nB(d * p - \epsilon_0), p(\mathbf{Y} | \mathbf{z}_0) \leq 2^{-B(H(Y^n | Z^n) - \epsilon_0)}, d(\mathbf{Y}, \mathbf{z}_0) \in [nB(p * p - \epsilon_0), nB(p * p + \epsilon_0)] | \mathbf{z}_0) \\ & \geq 1 - (\delta + 1 - q_d(1 - \delta)) \\ & = q_d(1 - \delta) - \delta \\ & \geq q_d - 2\delta. \end{aligned}$$

On the other hand,

$$\begin{aligned} & \Pr(d(\mathbf{Y}, \mathbf{y}_0) \geq nB(d * p - \epsilon_0), p(\mathbf{Y} | \mathbf{z}_0) \leq 2^{-B(H(Y^n | Z^n) - \epsilon_0)}, d(\mathbf{Y}, \mathbf{z}_0) \in [nB(p * p - \epsilon_0), nB(p * p + \epsilon_0)] | \mathbf{z}_0) \\ & \leq 2^{-B(H(Y^n | Z^n) - \epsilon_0)} |\{\mathbf{y} : d(\mathbf{y}, \mathbf{y}_0) \geq nB(d * p - \epsilon_0), d(\mathbf{y}, \mathbf{z}_0) \in [nB(p * p - \epsilon_0), nB(p * p + \epsilon_0)]\}| \\ & = 2^{-B(H(Y^n | Z^n) - \epsilon_0)} \left| \bigcup_{r=d*p-\epsilon_0}^1 \text{Sphere}(\mathbf{y}_0, nBr) \bigcap \bigcup_{\rho=p*p-\epsilon_0}^{p*p+\epsilon_0} \text{Sphere}(\mathbf{z}_0, nB\rho) \right| \\ & = 2^{-B(H(Y^n | Z^n) - \epsilon_0)} \left| \bigcup_{r=d*p-\epsilon_0}^1 \bigcup_{\rho=p*p-\epsilon_0}^{p*p+\epsilon_0} \underbrace{\text{Sphere}(\mathbf{y}_0, nBr) \bigcap \text{Sphere}(\mathbf{z}_0, nB\rho)}_{\text{Inter}(r, \rho)} \right|. \end{aligned}$$

Therefore,

$$q_d \leq 2\delta + 2^{-B(H(Y^n | Z^n) - \epsilon_0)} \left| \bigcup_{r=d*p-\epsilon_0}^1 \bigcup_{\rho=p*p-\epsilon_0}^{p*p+\epsilon_0} \text{Inter}(r, \rho) \right|. \quad (89)$$

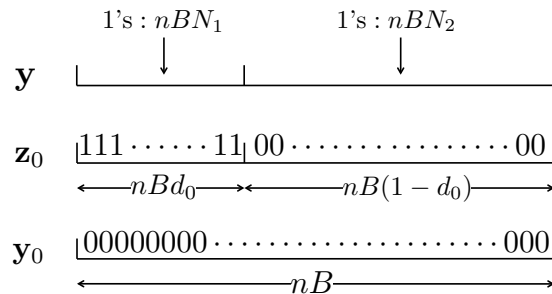


Fig. 6. Illustration of a specific pair $(\mathbf{y}_0, \mathbf{z}_0)$.

Now we show that the second term on the R.H.S. of (89) vanishes if $d > d_0 * p$. Without loss of generality, consider a specific pair $(\mathbf{y}_0, \mathbf{z}_0)$ as shown in Fig. 6, where $\mathbf{y}_0 = \mathbf{0}$ and $d(\mathbf{y}_0, \mathbf{z}_0) = nd_0$. We first characterize the volume of $\text{Inter}(r, \rho)$ for any r and $\rho = p * p$, i.e.,

$$\left| \text{Sphere}(\mathbf{y}_0, nBr) \cap \text{Sphere}(\mathbf{z}_0, nBp * p) \right|. \quad (90)$$

For any \mathbf{y} , let nBN_1 denote the number of 1's appearing in its first nBd_0 digits, and nBN_2 denote the number of 1's in the rest. Then the volume in (90) amounts to the number of \mathbf{y} 's such that the following two equalities hold

$$\begin{cases} N_1 + N_2 = r \\ d_0 - N_1 + N_2 = p * p \end{cases}$$

i.e.,

$$\begin{cases} N_1 = \frac{r + d_0 - p * p}{2} \\ N_2 = \frac{r + p * p - d_0}{2}. \end{cases}$$

Here, $N_1 \in [0, d_0]$ and $N_2 \in [0, (1 - d_0)]$, i.e.,

$$\begin{cases} r \in [p * p - d_0, p * p + d_0] \\ r \in [d_0 - p * p, 2 - p * p - d_0]. \end{cases} \quad (91)$$

$$(92)$$

Therefore, with $\rho = p * p$,

$$\begin{aligned} \text{Inter}(r, \rho) &= \binom{nBd_0}{nB\frac{r+d_0-p*p}{2}} \binom{nB(1-d_0)}{nB\frac{r+p*p-d_0}{2}} \\ &\leq 2^{nBd_0 \left(H\left(\frac{r+d_0-p*p}{2d_0}\right) \right)} 2^{nB(1-d_0) \left(H\left(\frac{r+p*p-d_0}{2(1-d_0)}\right) \right)} \\ &= 2^{nB \left(d_0 H\left(\frac{r+d_0-p*p}{2d_0}\right) + (1-d_0) H\left(\frac{r+p*p-d_0}{2(1-d_0)}\right) \right)} \end{aligned} \quad (93)$$

for sufficiently large B , where (93) follows from the bound $\binom{n}{nk} \leq \frac{1}{\sqrt{n\pi k(1-k)}} 2^{nH(k)}$ for any $k \in (0, 1)$, as stated in [29, Lemma 17.5.1].

Let $f(r) = d_0 H\left(\frac{r+d_0-p*p}{2d_0}\right) + (1-d_0) H\left(\frac{r+p*p-d_0}{2(1-d_0)}\right)$. It can be verified that $f(r)$ attains the maximum $H(p * p)$ if and only if $r = d_0 * p * p$; see Appendix L. Thus, when

$$\begin{aligned} \rho &= p * p \\ d &= d_0 * p + \epsilon' \\ r &\geq d * p - \epsilon_0 = (d_0 * p + \epsilon') * p - \epsilon_0 = d_0 * p * p + \epsilon'(1 - 2p) - \epsilon_0, \end{aligned}$$

we have

$$\text{Inter}(r, \rho) \leq 2^{nB(H(p*p) - \epsilon_1)}$$

for some $\epsilon_1 > 0$ provided $\epsilon'(1 - 2p) - \epsilon_0 > 0$. Further, due to the continuity of $\text{Inter}(r, \rho)$ in ρ , for any $\rho \in [p * p - \epsilon_0, p * p + \epsilon_0]$, $d = d_0 * p + \epsilon'$, $r \geq d * p - \epsilon_0$,

$$\text{Inter}(r, \rho) \leq 2^{nB(H(p*p) - \epsilon_1 + \epsilon_2)} \quad (94)$$

for some $\epsilon_2 \rightarrow 0$ as $\epsilon_0 \rightarrow 0$.

Plugging (94) into (89), we have for $d = d_0 * p + \epsilon'$ and sufficiently large B ,

$$\begin{aligned} q_d &\leq 2\delta + 2^{-B(H(Y^n|Z^n) - \epsilon_0)} \sum_{r=d*p-\epsilon_0}^1 \sum_{\rho=p*p-\epsilon_0}^{p*p+\epsilon_0} 2^{nB(H(p*p) - \epsilon_1 + \epsilon_2)} \\ &\leq 2\delta + 2^{-B(H(Y^n|Z^n) - \epsilon_0)} 2^{nB(H(p*p) - \epsilon_1 + \epsilon_0 + \epsilon_2)} \\ &\leq 2\delta + 2^{-nB(\frac{1}{n}H(Y^n|Z^n) - H(p*p) + \epsilon_1 - 2\epsilon_0 - \epsilon_2)}, \end{aligned}$$

with

$$\begin{aligned} \frac{1}{n}H(Y^n|Z^n) &= \frac{1}{n}(H(Y^n, Z^n) - H(Z^n)) \\ &= \frac{1}{n}(H(X^n) + H(Y^n, Z^n|X^n) - H(X^n|Y^n, Z^n) - H(Z^n)) \\ &= \frac{1}{n}(H(M) - H(M|X^n) + H(Y^n, Z^n|X^n) - H(X^n|Y^n, Z^n) - H(Z^n)) \\ &\geq \frac{1}{n}(nR - n\epsilon_0 + 2nH(p) - n\epsilon_0 - n) \\ &= C_{XYZ} - \tau + 2H(p) - 1 - 2\epsilon_0 \\ &= H(p * p) - \tau - 2\epsilon_0, \end{aligned} \tag{95}$$

for n sufficiently large, where in (95) we have used Fano's inequality. Thus, when τ, ϵ_0 are sufficiently small and n, B are sufficiently large, we have for any $\epsilon' > 0$, $d = d_0 * p + \epsilon'$,

$$\begin{aligned} q_d &\leq 2\delta + 2^{-nB(\epsilon_1 - \tau - 4\epsilon_0 - \epsilon_2)} \\ &\leq 3\delta. \end{aligned}$$

We finally conclude that for a $(\mathbf{y}_0, \mathbf{z}_0)$ pair where $\mathbf{z}_0 \in \mathcal{T}_\epsilon^{(B)}(Z^n)$ and $d(\mathbf{y}_0, \mathbf{z}_0) = nBd_0$, there exists some $\epsilon' \rightarrow 0$ as $\epsilon \rightarrow 0$ such that

$$\Pr(d(\mathbf{X}, \mathbf{y}_0) \leq nB(d_0 * p + \epsilon') | \mathbf{z}_0) \geq 1 - v.$$

where v can be made arbitrarily small by choosing n, B sufficiently large and τ sufficiently small.

IX. CONCLUSION

We consider the symmetric primitive relay channel, and develop two new upper bounds on its capacity that are tighter than existing bounds, including the celebrated cut-set bound. Our approach uses measure concentration (the blowing-up lemma in particular) to analyze the probabilistic geometric relations between the typical sets of the n -letter random variables associated with a reliable code for communicating over this channel. We then translate these relations to new entropy inequalities between the n -letter random variables involved.

Information theory and geometry are indeed known to be inherently related; for example the differential entropy of a continuous random variable can be regarded as the exponential growth rate of the volume of its typical set. Therefore, entropy relations can, in principle, be developed by studying the relative geometry of the typical sets of the random variables. However, we are not aware of many examples where such geometric techniques have been successfully used to develop converses for problems in network information theory. It would be interesting to see if the approach we develop in this paper, i.e. deriving information inequalities by studying the geometry of typical sets, in particular using measure concentration, can be used to make progress on other long-standing open problems in network information theory.

While we have exclusively focused on the symmetric relay channel in this paper, our results can be extended to asymmetric primitive relay channels [31] using the idea of channel simulation. An extension of these ideas to the Gaussian case has been provided in [32].

APPENDIX A

To see $E(R) \leq R - C_{XY}$ for any $R > I(X; Y)$, recall that $E(R)$ has the following alternative form [27]:

$$E(R) = \min_{p(x)} \min_{\tilde{p}(y|x)} D(\tilde{p}(y|x) || p(y|x)p(x)) + |R - I(p(x), \tilde{p}(y|x))|^+ \quad (96)$$

where $|t|^+ := \max\{0, t\}$, $D(\tilde{p}(y|x) || p(y|x)p(x))$ is the conditional relative entropy defined as

$$D(\tilde{p}(y|x) || p(y|x)p(x)) := \sum_{(x,y)} p(x)\tilde{p}(y|x) \log \frac{\tilde{p}(y|x)}{p(y|x)},$$

and $I(p(x), \tilde{p}(y|x))$ is the mutual information defined with respect to the joint distribution $p(x)\tilde{p}(y|x)$, i.e.,

$$I(p(x), \tilde{p}(y|x)) := \sum_{(x,y)} p(x)\tilde{p}(y|x) \log \frac{\tilde{p}(y|x)}{\sum_x p(x)\tilde{p}(y|x)}.$$

In the regime of $R > C_{XY}$, simply choosing the $p(x)$ and $\tilde{p}(y|x)$ in (96) to be capacity-achieving distribution $p^*(x)$ and $p(y|x)$ respectively would make the objective function equal to $R - C_{XY}$, and thus $E(R) \leq R - C_{XY}$.

APPENDIX B

We demonstrate the improvements of our bound in Theorem 4.1 over Xue's bound using the following simple example.

Example B.1: Suppose both X - Y and X - Z links are the binary asymmetric channels as depicted in Fig. 7, with parameters $p_1 = 0.01$ and $p_2 = 0.3$. For the input distribution

$$p(x) = \begin{cases} \alpha & x = 0 \\ 1 - \alpha & x = 1 \end{cases}$$

we have

$$I(X; Y) = H(\alpha(1 - p_1) + (1 - \alpha)p_2) - (\alpha H(p_1) + (1 - \alpha)H(p_2))$$

and

$$\begin{aligned} I(X; Y, Z) &= H([\alpha(1 - p_1)^2 + (1 - \alpha)p_2^2, \alpha(1 - p_1)p_1 + (1 - \alpha)(1 - p_2)p_2, \\ &\quad \alpha(1 - p_1)p_1 + (1 - \alpha)(1 - p_2)p_2, \alpha p_1^2 + (1 - \alpha)(1 - p_2)^2]) \\ &\quad - 2(\alpha H(p_1) + (1 - \alpha)H(p_2)). \end{aligned}$$

With $p_1 = 0.01$ and $p_2 = 0.3$, numerical evaluation of $I(X; Y)$ and $I(X; Y, Z)$ yields that $C_{XY} = \max_{\alpha} I(X; Y) = 0.46432$ with the maximizer $\alpha_{XY}^* = 0.58$, and $C_{XYZ} = \max_{\alpha} I(X; Y, Z) = 0.72022$ with the maximizer $\alpha_{XYZ}^* = 0.54$.

Suppose we want to achieve a rate $R = C_{XYZ}$, and we use Proposition 3.3 and Theorem 4.1 to derive a lower bound on R_0 , respectively. First consider Proposition 3.3. Numerically, we have $E(R) = 0.05951$ for $R = C_{XYZ} = 0.72022$, and the minimum a to satisfy (6) is $a = 0.00008$. Thus, by (5), we have

$$\begin{aligned} R_0 &\geq R - C_{XY} + a \\ &\geq 0.72022 - 0.46432 + 0.00008 \\ &= 0.25598. \end{aligned} \quad (97)$$

We then apply Theorem 4.1 and demonstrate that the improvements mentioned in Section IV-B result in a tighter bound on R_0 . In Theorem 4.1, $p(x)$ has to be chosen such that $\alpha = \alpha_{XYZ}^* = 0.54$ due to the constraint (14). Under such a distribution of $p(x)$, numerically, we have $I(X; Y) = 0.46223 < C_{XY}$, and

$C_{XYZ} - I(X; Y) = 0.25799 > E(R)$. Noting the R.H.S. of (6) is also sharpened to that of (17), we can calculate the minimum a satisfying (17) to be $a = 0.00546$. Thus, by (15), we have

$$\begin{aligned} R_0 &\geq R - I(X; Y) + a \\ &\geq 0.72022 - 0.46223 + 0.00546 \\ &= 0.26345, \end{aligned} \quad (98)$$

where it is easy to see that the last two terms in (98) are both sharpened compared to those in (97).

Therefore, in order to achieve the rate $R = C_{XYZ}$, the lower bounds on R_0 yielded by Proposition 3.3 and Theorem 4.1 are

$$R_0 \geq 0.25598$$

and

$$R_0 \geq 0.26345$$

respectively. Viewed from another perspective, for $R_0 \in [0.25598, 0.26345)$, the bound in Theorem 4.1 asserts that the capacity of the relay channel $C(R_0) < C_{XYZ}$, which excludes the possibility of achieving $R = C_{XYZ}$ while Xue's bound in Proposition 3.3 cannot.

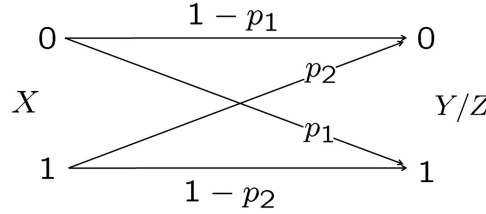


Fig. 7. Binary asymmetric channel.

APPENDIX C $\Delta(p(x), d)$ FOR BINARY SYMMETRIC CHANNELS

For a binary symmetric channel with crossover probability $p < 1/2$, the objective function in (18) can be expressed as

$$\begin{aligned} &H(\tilde{p}(\omega|x)|p(x)) + D(\tilde{p}(\omega|x)||p(\omega|x)|p(x)) - H(p(\omega|x)|p(x)) \\ &= - \sum_{(x,\omega)} p(x)\tilde{p}(\omega|x) \log \tilde{p}(\omega|x) + \sum_{(x,\omega)} p(x)\tilde{p}(\omega|x) \log \frac{\tilde{p}(\omega|x)}{p(\omega|x)} - \sum_{(x,\omega)} p(x)p(\omega|x) \log \frac{1}{p(\omega|x)} \\ &= \sum_{(x,\omega)} [p(x)\tilde{p}(\omega|x) - p(x)p(\omega|x)] \log \frac{1}{p(\omega|x)} \\ &= \sum_{x \neq \omega} [p(x)\tilde{p}(\omega|x) - p(x)p(\omega|x)] \log \frac{1}{p} + \sum_{x=\omega} [p(x)\tilde{p}(\omega|x) - p(x)p(\omega|x)] \log \frac{1}{1-p} \end{aligned} \quad (99)$$

$$= \sum_{x \neq \omega} [p(x)\tilde{p}(\omega|x) - p(x)p(\omega|x)] \log \frac{1}{p} + \sum_{x \neq \omega} [p(x)(1 - \tilde{p}(\omega|x)) - p(x)(1 - p(\omega|x))] \log \frac{1}{1-p} \quad (100)$$

$$= \sum_{x \neq \omega} [p(x)\tilde{p}(\omega|x) - p(x)p(\omega|x)] \log \frac{1}{p} - \sum_{x \neq \omega} [p(x)\tilde{p}(\omega|x) - p(x)p(\omega|x)] \log \frac{1}{1-p} \quad (101)$$

$$= \sum_{x \neq \omega} [p(x)\tilde{p}(\omega|x) - p(x)p(\omega|x)] \log \frac{1-p}{p} \quad (102)$$

We now show that under the constraint

$$\frac{1}{2} \sum_{(x,\omega)} |p(x)\tilde{p}(\omega|x) - p(x)p(\omega|x)| \leq d, \quad (103)$$

the function in (102), and thus $\Delta(p(x), d)$, are upper bounded by

$$\min \{d, 1 - p\} \log \frac{1 - p}{p}.$$

Along the similar lines as in (99)–(101), we obtain

$$\sum_{x=\omega} |p(x)\tilde{p}(\omega|x) - p(x)p(\omega|x)| = \sum_{x \neq \omega} |p(x)\tilde{p}(\omega|x) - p(x)p(\omega|x)|,$$

and thus the constraint (103) can be rewritten as

$$\sum_{x \neq \omega} |p(x)\tilde{p}(\omega|x) - p(x)p(\omega|x)| \leq d. \quad (104)$$

On the other hand, we have

$$\begin{aligned} \sum_{x \neq \omega} |p(x)\tilde{p}(\omega|x) - p(x)p(\omega|x)| &= \sum_{x \neq \omega} p(x) |\tilde{p}(\omega|x) - p(\omega|x)| \\ &\leq \sum_{x \neq \omega} p(x)(1 - p) \\ &= 1 - p. \end{aligned} \quad (105)$$

Combining (102), (104) and (105) yields that

$$\Delta(p(x), d) \leq \min \{d, 1 - p\} \log \frac{1 - p}{p}.$$

In fact, it can be easily checked that the equality sign in the above inequality can be attained by choosing

$$\tilde{p}(\omega|x) = p + \min \{d, 1 - p\}, \forall (x, \omega) \text{ with } x \neq \omega,$$

and thus we conclude that

$$\Delta(p(x), d) = \min \{d, 1 - p\} \log \frac{1 - p}{p}.$$

APPENDIX D

UPPER BOUNDS FOR BINARY SYMMETRIC CHANNEL CASE

Various upper bounds are evaluated for the binary symmetric channel case as follows.

A. Cut-Set bound (Prop. 3.1)

The optimal distribution for Prop. 3.1 is $p^*(0) = p^*(1) = 1/2$, under which,

$$\begin{cases} I^*(X; Y, Z) = C_{XYZ} = 1 + H(p * p) - 2H(p) \\ I^*(X; Y) = C_{XY} = 1 - H(p). \end{cases}$$

Therefore, the cut-set bound simplifies to

$$C(R_0) \leq \min \{1 + H(p * p) - 2H(p), 1 - H(p) + R_0\}.$$

B. Xue's bound (Prop. 3.3)

Since the function $E(R)$ is monotonic in R , its inverse function $E^{-1}(\cdot)$ exists and Xue's bound can be expressed as

$$C(R_0) \leq \max_{a \in [0, R_0]} \min \left\{ 1 - H(p) + R_0 - a, E^{-1}(H(\sqrt{a}) + \sqrt{a}) \right\}.$$

C. Our first bound (Thm. 4.1)

The uniform distribution of X is also optimal for Thm. 4.1, under which our first bound reduces to

$$C(R_0) \leq \max_{a \in [0, \min\{R_0, H(p), \frac{1}{2 \ln 2}\}]} \min \left\{ 1 + H(p * p) - 2H(p), 1 - H(p) + R_0 - a, \right. \\ \left. 1 - H(p) + H\left(\sqrt{\frac{a \ln 2}{2}}\right) - a \right\}.$$

where the constraint of a follows since $H(Z|X) = H(p)$ for any $p(x)$ and $\frac{2}{\ln 2} \left(\frac{|\Omega|-1}{|\Omega|}\right)^2 = \frac{1}{2 \ln 2}$, and the term $\sqrt{\frac{a \ln 2}{2}} \log(|\Omega| - 1)$ disappears compared to Thm. 4.1 since it becomes 0 with $|\Omega| = 2$.

D. Our second bound (Thm. 4.2)

Recall that in the binary symmetric channel case, $\Delta(p(x), d)$ is independent of $p(x)$ and given by

$$\min\{d, 1 - p\} \log \frac{1 - p}{p} := \bar{\Delta}(d).$$

Therefore, our new bound becomes

$$C(R_0) \leq \max_{a \in [0, \min\{R_0, H(p)\}]} \min \left\{ 1 + H(p * p) - 2H(p), 1 - H(p) + R_0 - a, 1 - H(p) + \bar{\Delta}\left(\sqrt{\frac{a \ln 2}{2}}\right) \right\}. \quad (106)$$

It is not difficult to see that for the R.H.S. of (106), at least one of the maximizers must be no greater than $\frac{2}{\ln 2}(1 - p)^2$, i.e., satisfying $\sqrt{\frac{a \ln 2}{2}} \leq 1 - p$. Therefore, (106) can be equivalently stated as

$$C(R_0) \leq \max_{a \in [0, \min\{R_0, H(p), \frac{2}{\ln 2}(1-p)^2\}]} \min \left\{ 1 + H(p * p) - 2H(p), 1 - H(p) + R_0 - a, \right. \\ \left. 1 - H(p) + \sqrt{\frac{a \ln 2}{2}} \log \frac{1 - p}{p} \right\}.$$

APPENDIX E
PROOF OF (26)

To show (26), it suffices to show $H(p)/H(p * p) \rightarrow 1/2$ as $p \rightarrow 0$. For this, we have

$$\begin{aligned}
\lim_{p \rightarrow 0} \frac{H(p)}{H(p * p)} &= \lim_{p \rightarrow 0} \frac{H'(p)}{H'(p * p) \cdot (p * p)'} \\
&= \lim_{p \rightarrow 0} \frac{\log \frac{1-p}{p}}{\log \frac{1-p*p}{p*p} \cdot (2-4p)} \\
&= \frac{1}{2} \lim_{p \rightarrow 0} \frac{\log' \frac{1-p}{p} \cdot (\frac{1-p}{p})'}{\log' \frac{1-p*p}{p*p} \cdot (\frac{1-p*p}{p*p})'} \\
&= \frac{1}{2} \lim_{p \rightarrow 0} \frac{\frac{p}{1-p} \cdot (-\frac{1}{p^2})}{\frac{p*p}{1-p*p} \cdot (-\frac{1}{(p*p)^2}) \cdot (2-4p)} \\
&= \frac{1}{4} \lim_{p \rightarrow 0} \frac{(p * p)(1 - p * p)}{p(1 - p)} \\
&= \frac{1}{4} \lim_{p \rightarrow 0} 2(1 - p * p) \\
&= \frac{1}{2}
\end{aligned}$$

which proves (26).

APPENDIX F
PROOF OF LEMMA 6.1

To prove Lemma 6.1, we need the following lemma, whose proof relies on the property of polynomial number of compositions and can be easily extended from the proof for a single user channel [24].

Lemma F.1: Suppose a code $(\mathcal{C}_{(n,R)}, f_n, g_n)$ has the probability of error $P_e^{(n)}$. Then there is some composition Q for which a fixed composition code $(\mathcal{C}_{(n,\tilde{R})}^{[Q]}, f_n, g_n)$ exists, and its probability of error $\tilde{P}_e^{(n)}$ and rate \tilde{R} satisfy $\tilde{P}_e^{(n)} \leq P_e^{(n)}$ and $\tilde{R} \geq R - (|\Omega_X| - 1) \frac{\log(n+1)}{n}$.

Proof of Lemma 6.1: Since R is achievable, there exists a sequence of codes

$$\{(\mathcal{C}_{(n,R)}, f_n, g_n)\}_{n=1}^{\infty}$$

such that $P_e^{(n)} \rightarrow 0$ as $n \rightarrow \infty$. By Lemma F.1, there accordingly exists a sequence of fixed composition codes

$$\{(\mathcal{C}_{(n,\tilde{R})}^{[Q_n]}, f_n, g_n)\}_{n=1}^{\infty}$$

such that the probability of error $\tilde{P}_e^{(n)} \leq P_e^{(n)}$ and the rate $\tilde{R} \geq R - (|\Omega_X| - 1) \frac{\log(n+1)}{n}$. Noting that $\tilde{R} \rightarrow R$ and $\tilde{P}_e^{(n)} \rightarrow 0$ as $n \rightarrow \infty$, Lemma 6.1 is thus proved. ■

APPENDIX G
PROOF OF LEMMA 6.2

We only characterize $H(Y^n|X^n)$ and $I(X^n; Y^n)$ while the other information quantities can be characterized similarly. Consider $H(Y^n|x^n)$ for any specific x^n with composition Q_n . We have

$$\begin{aligned} H(Y^n|x^n) &= \sum_{i=1}^n H(Y_i|x^n, Y^{i-1}) \\ &= \sum_{i=1}^n H(Y_i|x_i) \\ &= \sum_x nQ_n(x)H(Y|x) \\ &= nH(Y|X) \end{aligned}$$

where $H(Y|X)$ is calculated based on $Q_n(x)p(\omega|x)$. Therefore, for the code with fixed composition Q_n ,

$$H(Y^n|X^n) = \sum_{x^n} p(x^n)H(Y^n|x^n) = \sum_{x^n} p(x^n)(nH(Y|X)) = nH(Y|X).$$

To bound $I(X^n; Y^n)$, it suffices to bound $H(Y^n)$. For any specific x^n with composition Q_n , we have for some $\epsilon_1 \rightarrow 0$ as $n \rightarrow \infty$,

$$\Pr(Y^n \in \mathcal{T}_{\epsilon_1}^{(n)}(Y)|x^n) \geq 1 - \epsilon_1,$$

where $\mathcal{T}_{\epsilon_1}^{(n)}(Y)$ is the typical set [30] with respect to $\sum_x Q_n(x)p(\omega|x)$. Therefore, for the code with fixed composition Q_n ,

$$\Pr(Y^n \in \mathcal{T}_{\epsilon_1}^{(n)}(Y)) = \sum_{x^n} p(x^n)\Pr(Y^n \in \mathcal{T}_{\epsilon_1}^{(n)}(Y)|x^n) \geq 1 - \epsilon_1.$$

Letting $W = \mathbb{I}(Y^n \in \mathcal{T}_{\epsilon_1}^{(n)}(Y))$, we have

$$\begin{aligned} H(Y^n) &\leq H(Y^n, W) \\ &\leq 1 + H(Y^n|W) \\ &= 1 + \Pr(Y^n \in \mathcal{T}_{\epsilon_1}^{(n)}(Y))H(Y^n|Y^n \in \mathcal{T}_{\epsilon_1}^{(n)}(Y)) + \Pr(Y^n \notin \mathcal{T}_{\epsilon_1}^{(n)}(Y))H(Y^n|Y^n \notin \mathcal{T}_{\epsilon_1}^{(n)}(Y)) \\ &\leq 1 + \log |\mathcal{T}_{\epsilon_1}^{(n)}(Y)| + n\epsilon_1 \log |\Omega| \\ &\leq 1 + n(H(Y) + \epsilon_2) + n\epsilon_1 \log |\Omega| \\ &\leq n(H(Y) + \epsilon) \end{aligned}$$

where $\epsilon_1, \epsilon_2, \epsilon \rightarrow 0$ as $n \rightarrow \infty$, and $H(Y)$ is calculated based on $\sum_x Q_n(x)p(\omega|x)$. Combining this with the fact that $H(Y^n|X^n) = nH(Y|X)$, we have

$$I(X^n; Y^n) \leq n(I(X; Y) + \epsilon). \quad (107)$$

We now argue that the ϵ in (107) can be dropped. Given any fixed n , consider a length- B sequence of i.i.d. random vector pairs $\{(X^n(b), Y^n(b))\}_{b=1}^B$, denoted by (\mathbf{X}, \mathbf{Y}) , where $(X^n(b), Y^n(b))$ have the same distribution as (X^n, Y^n) for any $b \in [1 : B]$. Obviously the length- nB vector \mathbf{X} also has composition Q_n , and by (107) we have

$$I(\mathbf{X}; \mathbf{Y}) \leq nB(I(X; Y) + \epsilon),$$

where $\epsilon \rightarrow 0$ as $B \rightarrow \infty$. Due to the i.i.d. property, we further have

$$BI(X^n; Y^n) \leq nB(I(X; Y) + \epsilon).$$

Dividing B at both sides of the above equation and letting $B \rightarrow \infty$, we obtain $I(X^n; Y^n) \leq nI(X; Y)$.

APPENDIX H

VOLUME OF A HAMMING BALL

Consider the volume of a general n -dimensional Hamming ball in Ω^n with radius nr . It is obvious that when $r \geq 1$, the volume

$$|\text{Ball}(nr)| = |\Omega|^n = 2^{n \log |\Omega|}. \quad (108)$$

For $r < 1$, we have

$$\begin{aligned} |\text{Ball}(nr)| &= 1 + \sum_{k=\frac{1}{n}}^r \binom{n}{nk} (|\Omega| - 1)^{nk} \\ &\leq 1 + \sum_{k=\frac{1}{n}}^r \frac{2^{nH(k)}}{\sqrt{\pi nk(1-k)}} (|\Omega| - 1)^{nk} \\ &\leq \sum_{k=0}^r 2^{n(H(k)+k \log(|\Omega|-1))} \\ &\leq (nr + 1) \max_{k \in \{0, \frac{1}{n}, \dots, r\}} 2^{n(H(k)+k \log(|\Omega|-1))} \\ &\leq \max_{k \in \{0, \frac{1}{n}, \dots, r\}} 2^{n(H(k)+k \log(|\Omega|-1)+\epsilon)} \\ &= 2^{n \left(\max_{k \in \{0, \frac{1}{n}, \dots, r\}} H(k) + k \log(|\Omega|-1) + \epsilon \right)} \end{aligned}$$

for any $\epsilon > 0$ and sufficiently large n , where the first inequality follows from [29, Lemma 17.5.1]. Similarly, we can lower bound $|\text{Ball}(nr)|$ as

$$|\text{Ball}(nr)| \geq 2^{n \left(\max_{k \in \{0, \frac{1}{n}, \dots, r\}} H(k) + k \log(|\Omega|-1) - \epsilon \right)}$$

for any $\epsilon > 0$ and sufficiently large n . Therefore, we have for $r < 1$,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log |\text{Ball}(nr)| = \max_{t \in [0, r]} H(t) + t \log(|\Omega| - 1). \quad (109)$$

Now we simplify the above expression. Let $v(t) = H(t) + t \log(|\Omega| - 1)$ for $t \in (0, 1)$. We have $v'(t) = \log \frac{(1-t)(|\Omega|-1)}{t}$, which is decreasing in t for $t \in (0, 1)$ and equals 0 when $t = \frac{|\Omega|-1}{|\Omega|}$. Thus, the maximum of $v(t)$ is attained when $t = \frac{|\Omega|-1}{|\Omega|}$, and is given by

$$\begin{aligned} v^*(t) &= H\left(\frac{|\Omega|-1}{|\Omega|}\right) + \frac{|\Omega|-1}{|\Omega|} \log(|\Omega| - 1) \\ &= -\frac{|\Omega|-1}{|\Omega|} \log \frac{|\Omega|-1}{|\Omega|} - \frac{1}{|\Omega|} \log \frac{1}{|\Omega|} + \frac{|\Omega|-1}{|\Omega|} \log(|\Omega| - 1) \\ &= \frac{|\Omega|-1}{|\Omega|} \log \frac{(|\Omega|-1)|\Omega|}{(|\Omega|-1)} - \frac{1}{|\Omega|} \log \frac{1}{|\Omega|} \\ &= \frac{|\Omega|-1}{|\Omega|} \log |\Omega| + \frac{1}{|\Omega|} \log |\Omega| \\ &= \log |\Omega|. \end{aligned} \quad (110)$$

Therefore, we have

$$\max_{t \in [0, r]} H(t) + t \log(|\Omega| - 1) = \begin{cases} \log |\Omega| & \text{when } r \in \left(\frac{|\Omega|-1}{|\Omega|}, 1\right) \\ H(r) + r \log(|\Omega| - 1) & \text{when } r \in \left[0, \frac{|\Omega|-1}{|\Omega|}\right]. \end{cases} \quad (111)$$

$$(112)$$

Combining (108), (109), (111) and (112), we obtain that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log |\text{Ball}(nr)| = \begin{cases} \log |\Omega| & \text{when } r > \frac{|\Omega|-1}{|\Omega|} \\ H(r) + r \log(|\Omega| - 1) & \text{when } r \leq \frac{|\Omega|-1}{|\Omega|}. \end{cases}$$

APPENDIX I

For any $(\mathbf{x}, \mathbf{z}) \in \mathcal{T}_\epsilon^{(B)}(X^n, Z^n)$, we have

$$\begin{aligned} |P_{\mathbf{x}}(x^n) - p(x^n)| &\leq \epsilon p(x^n), \forall x^n \\ |P_{(\mathbf{x}, \mathbf{z})}(x^n, z^n) - p(x^n, z^n)| &\leq \epsilon p(x^n, z^n), \forall (x^n, z^n) \end{aligned}$$

and thus

$$|P_{\mathbf{z}|\mathbf{x}}(z^n|x^n) - p(z^n|x^n)| \leq \epsilon_1 p(z^n|x^n), \forall (x^n, z^n) \text{ with } P_{\mathbf{x}}(x^n) \neq 0, \quad (113)$$

for some $\epsilon_1 \rightarrow 0$ as $\epsilon \rightarrow 0$.

Therefore, we have for any (x, ω) that

$$\begin{aligned} P_{(\mathbf{x}, \mathbf{z})}(x, \omega) &= \frac{1}{nB} N(x, \omega | \mathbf{x}, \mathbf{z}) \\ &= \frac{1}{nB} \sum_{(x^n, z^n)} N(x^n, z^n | \mathbf{x}, \mathbf{z}) \cdot N(x, \omega | x^n, z^n) \\ &= \sum_{(x^n, z^n)} P_{(\mathbf{x}, \mathbf{z})}(x^n, z^n) \cdot P_{(x^n, z^n)}(x, \omega) \\ &= \sum_{x^n: P_{\mathbf{x}}(x^n) > 0} P_{\mathbf{x}}(x^n) \sum_{z^n} P_{\mathbf{z}|\mathbf{x}}(z^n|x^n) \cdot P_{(x^n, z^n)}(x, \omega) \\ &\leq \sum_{x^n: P_{\mathbf{x}}(x^n) > 0} P_{\mathbf{x}}(x^n) \sum_{z^n} p(z^n|x^n)(1 + \epsilon_1) \cdot P_{(x^n, z^n)}(x, \omega) \\ &= (1 + \epsilon_1) \sum_{x^n: P_{\mathbf{x}}(x^n) > 0} P_{\mathbf{x}}(x^n) E[P_{(x^n, Z^n)}(x, \omega)] \\ &= (1 + \epsilon_1) \sum_{x^n: P_{\mathbf{x}}(x^n) > 0} P_{\mathbf{x}}(x^n) E \left[\frac{1}{n} N(x, \omega | x^n, Z^n) \right] \\ &= (1 + \epsilon_1) \sum_{x^n: P_{\mathbf{x}}(x^n) > 0} P_{\mathbf{x}}(x^n) E \left[\frac{1}{n} \sum_{j: x_j = x} \mathbb{I}(Z_j = \omega) \right] \\ &= (1 + \epsilon_1) \sum_{x^n: P_{\mathbf{x}}(x^n) > 0} P_{\mathbf{x}}(x^n) P_{x^n}(x) p(\omega|x) \\ &= (1 + \epsilon_1) P_{\mathbf{x}}(x) p(\omega|x) \\ &= (1 + \epsilon_1) Q_n(x) p(\omega|x). \end{aligned}$$

Similarly,

$$P_{(\mathbf{x}, \mathbf{z})}(x, \omega) \geq (1 - \epsilon_1) Q_n(x) p(\omega|x), \forall (x, \omega),$$

and thus

$$|P_{(\mathbf{x}, \mathbf{z})}(x, \omega) - Q_n(x) p(\omega|x)| \leq \epsilon_1 Q_n(x) p(\omega|x), \forall (x, \omega).$$

APPENDIX J

For any $(\mathbf{x}, \mathbf{y}, \mathbf{z})$, consider the total variation distance between $P_{(\mathbf{x}, \mathbf{y})}(x, \omega)$ and $P_{(\mathbf{x}, \mathbf{z})}(x, \omega)$. We have

$$\begin{aligned}
& nB \sum_{(x, \omega)} |P_{(\mathbf{x}, \mathbf{y})}(x, \omega) - P_{(\mathbf{x}, \mathbf{z})}(x, \omega)| \\
&= \sum_{(x, \omega)} \left| \sum_{i=1}^{nB} \mathbb{I}((x_i, y_i) = (x, \omega)) - \mathbb{I}((x_i, z_i) = (x, \omega)) \right| \\
&= \sum_{(x, \omega)} \left| \sum_{i: y_i = z_i} \mathbb{I}((x_i, y_i) = (x, \omega)) - \mathbb{I}((x_i, z_i) = (x, \omega)) + \sum_{i: y_i \neq z_i} \mathbb{I}((x_i, y_i) = (x, \omega)) - \mathbb{I}((x_i, z_i) = (x, \omega)) \right| \\
&= \sum_{(x, \omega)} \left| \sum_{i: y_i \neq z_i} \mathbb{I}((x_i, y_i) = (x, \omega)) - \mathbb{I}((x_i, z_i) = (x, \omega)) \right| \\
&\leq \sum_{(x, \omega)} \sum_{i: y_i \neq z_i} \mathbb{I}((x_i, y_i) = (x, \omega)) + \sum_{(x, \omega)} \sum_{i: y_i \neq z_i} \mathbb{I}((x_i, z_i) = (x, \omega)) \\
&= 2d(\mathbf{y}, \mathbf{z}),
\end{aligned}$$

i.e.,

$$\sum_{(x, \omega)} |P_{(\mathbf{x}, \mathbf{y})}(x, \omega) - P_{(\mathbf{x}, \mathbf{z})}(x, \omega)| \leq \frac{2}{nB} d(\mathbf{y}, \mathbf{z}).$$

APPENDIX K

From the property of jointly typical sequences, for any $\mathbf{z}_0 \in \mathcal{T}_\epsilon^{(B)}(Z^n)$ and $\epsilon_1 > \epsilon$,

$$\Pr((\mathbf{X}, \mathbf{Y}, \mathbf{z}_0) \in \mathcal{T}_{\epsilon_1}^{(B)}(X^n, Y^n, Z^n) | \mathbf{z}_0) \rightarrow 1 \text{ as } B \rightarrow \infty. \quad (114)$$

For any $(\mathbf{y}, \mathbf{z}_0) \in \mathcal{T}_{\epsilon_1}^{(B)}(Y^n, Z^n)$,

$$p(\mathbf{y} | \mathbf{z}_0) \leq 2^{-B(H(Y^n | Z^n) - \epsilon_2)}, \text{ for some } \epsilon_2 \rightarrow 0 \text{ as } \epsilon_1 \rightarrow 0. \quad (115)$$

Also, along the same lines as Appendix I, it can be shown that if $(\mathbf{x}, \mathbf{y}, \mathbf{z}_0)$ are jointly typical with respect to the n -letter random variables (X^n, Y^n, Z^n) , then $(\mathbf{x}, \mathbf{y}, \mathbf{z}_0)$ are also jointly typical with respect to the single-letter random variables (X, Y, Z) , i.e.,

$$|P_{(\mathbf{x}, \mathbf{y}, \mathbf{z}_0)}(x, y, z) - P_{\mathbf{x}}(x)p(y|x)p(z|x)| \leq \epsilon_3 P_{\mathbf{x}}(x)p(y|x)p(z|x),$$

for some $\epsilon_3 \rightarrow 0$ as $\epsilon_1 \rightarrow 0$. Then,

$$\begin{aligned}
P_{(\mathbf{y}, \mathbf{z}_0)}(0, 1) &\leq P_{\mathbf{x}}(0)(1 - p)p(1 + \epsilon_3) + P_{\mathbf{x}}(1)p(1 - p)(1 + \epsilon_3) \\
&= p(1 - p)(1 + \epsilon_3)
\end{aligned}$$

and $P_{(\mathbf{y}, \mathbf{z}_0)}(0, 1) \geq p(1 - p)(1 - \epsilon_3)$. Similarly, we also have

$$P_{(\mathbf{y}, \mathbf{z}_0)}(1, 0) \in [p(1 - p)(1 - \epsilon_3), p(1 - p)(1 + \epsilon_3)],$$

and thus

$$\begin{aligned}
d(\mathbf{y}, \mathbf{z}_0) &= nBP_{(\mathbf{y}, \mathbf{z}_0)}(0, 1) + nBP_{(\mathbf{y}, \mathbf{z}_0)}(1, 0) \\
&\in [2nBp(1 - p)(1 - \epsilon_3), 2nBp(1 - p)(1 + \epsilon_3)],
\end{aligned}$$

i.e.,

$$d(\mathbf{y}, \mathbf{z}_0) \in [nB(p * p - \epsilon_4), nB(p * p + \epsilon_4)], \quad (116)$$

where $\epsilon_4 \rightarrow 0$ as $\epsilon_1 \rightarrow 0$.

Combining (114), (115) and (116), we conclude that for any $\mathbf{z}_0 \in \mathcal{T}_\epsilon^{(B)}(Z^n)$ and some $\epsilon_0 \rightarrow 0$ as $\epsilon \rightarrow 0$,

$$\Pr(p(\mathbf{Y} | \mathbf{z}_0) \leq 2^{-B(H(Y^n | Z^n) - \epsilon_0)}, d(\mathbf{Y}, \mathbf{z}_0) \in [nB(p * p - \epsilon_0), nB(p * p + \epsilon_0)] | \mathbf{z}_0) \rightarrow 1 \text{ as } B \rightarrow \infty.$$

APPENDIX L
PROPERTY OF $f(r)$

For notational convenience, let $q := p * p$. Taking the first derivative of $f(r)$, we have

$$\begin{aligned} f'(r) &= d_0 H' \left(\frac{r + d_0 - q}{2d_0} \right) \cdot \left(\frac{r + d_0 - q}{2d_0} \right)' + (1 - d_0) H' \left(\frac{r + q - d_0}{2(1 - d_0)} \right) \cdot \left(\frac{r + q - d_0}{2(1 - d_0)} \right)' \\ &= \frac{1}{2} \left[\log \frac{d_0 - r + q}{r + d_0 - q} + \log \frac{2 - d_0 - r - q}{r + q - d_0} \right]. \end{aligned}$$

With $r = d_0 * q$, we have

$$\begin{aligned} f(d_0 * q) &= d_0 H \left(\frac{d_0 * q + d_0 - q}{2d_0} \right) + (1 - d_0) H \left(\frac{d_0 * q + q - d_0}{2(1 - d_0)} \right) \\ &= d_0 H \left(\frac{(d_0(1 - q) + (1 - d_0)q) + d_0 - q}{2d_0} \right) + (1 - d_0) H \left(\frac{(d_0(1 - q) + (1 - d_0)q) + q - d_0}{2(1 - d_0)} \right) \\ &= d_0 H(1 - q) + (1 - d_0) H(q) \\ &= H(q) \end{aligned}$$

and

$$\begin{aligned} f'(d_0 * q) &= \frac{1}{2} \left[\log \frac{d_0 - d_0 * q + q}{d_0 * q + d_0 - q} + \log \frac{2 - d_0 - d_0 * q - q}{d_0 * q + q - d_0} \right] \\ &= \frac{1}{2} \log \left[\frac{d_0 - (d_0(1 - q) + (1 - d_0)q) + q}{(d_0(1 - q) + (1 - d_0)q) + d_0 - q} \cdot \frac{2 - d_0 - (d_0(1 - q) + (1 - d_0)q) - q}{(d_0(1 - q) + (1 - d_0)q) + q - d_0} \right] \\ &= \frac{1}{2} \log \left[\frac{2d_0q}{2d_0(1 - q)} \cdot \frac{2(1 - d_0)(1 - q)}{2q(1 - d_0)} \right] \\ &= 0. \end{aligned}$$

Further taking the second derivative of $f(r)$ yields that

$$\begin{aligned} f''(r) &= \frac{1}{2 \ln 2} \left[\frac{r + d_0 - q}{d_0 - r + q} \left(\frac{d_0 - r + q}{r + d_0 - q} \right)' + \frac{r + q - d_0}{2 - d_0 - r - q} \left(\frac{2 - d_0 - r - q}{r + q - d_0} \right)' \right] \\ &= \frac{1}{2 \ln 2} \left[\frac{2d_0}{(r + d_0 - q)(r - d_0 - q)} + \frac{2(d_0 - 1)}{(2 - d_0 - r - q)(r - d_0 + q)} \right]. \end{aligned}$$

Provided the following constraint on r (cf. (91)–(92))

$$r \in (\max\{q - d_0, d_0 - q\}, \min\{q + d_0, 2 - q - d_0\})$$

it can be easily seen that $f''(r) < 0$. Therefore, $f(r)$ attains the maximum $H(p * p)$ if and only if $r = d_0 * q = d_0 * p * p$.

REFERENCES

- [1] X. Wu, L.-L. Xie, A. Ozgur, "Upper bounds on the capacity of symmetric primitive relay channels," in *Proc. of IEEE International Symposium on Information Theory 2015*, Hong Kong, June 14–19, 2015.
- [2] X. Wu and A. Ozgur, "Improving on the cut-set bound via geometric analysis of typical sets," in *Proc. of 2016 International Zurich Seminar on Communications*.
- [3] E. C. van der Meulen, "Three-terminal communication channels," *Adv. Appl. Prob.*, vol. 3, pp. 120–154, 1971.
- [4] T. Cover and A. El Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inform. Theory*, vol. 25, pp. 572–584, 1979.
- [5] L.-L. Xie and P. R. Kumar, "An achievable rate for the multiple-level relay channel," *IEEE Trans. Inform. Theory*, vol. 51, pp. 1348–1358, April 2005.
- [6] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Trans. Inform. Theory*, vol. 51, pp. 3037–3063, September 2005.

- [7] X. Wu and L.-L. Xie, "A unified relay framework with both D-F and C-F relay nodes," *IEEE Trans. Inform. Theory*, vol. 60, no. 1, pp. 586–604, January 2014.
- [8] B. Schein and R. Gallager, "The Gaussian parallel relay network," in *Proc. of IEEE International Symposium on Informaiton Theory*, pp. 22, June 2000.
- [9] A. S. Avestimehr, S. N. Diggavi, and D. N. C. Tse, "Wireless Network Information Flow: A Deterministic Approach," *IEEE Trans. Info. Theory*, vol. 57, no. 4, pp. 1872–1905, 2011.
- [10] B. Nazer and M. Gastpar, "Compute-and-forward: Harnessing interference through structured codes," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6463–6486, 2011.
- [11] S. H. Lim, Y.-H. Kim, A. El Gamal, S.-Y. Chung, "Noisy network coding," *IEEE Trans. Inform. Theory*, vol. 57, no. 5, pp. 3132–3152, May 2011.
- [12] P. Minero, S. H. Lim, and Y. Kim, "A unified approach to hybrid coding," *IEEE Trans. Inform. Theory*, vol. 61, no. 4, pp. 1509–1523, Apr. 2015.
- [13] S. Zahedi, "On reliable communication over relay channels," Ph.D. dissertation, Stanford Univ., Stanford, CA, 2005.
- [14] A. El Gamal and M. Aref, "The capacity of the semideterministic relay channel," *IEEE Trans. Inform. Theory*, vol. 28, no. 3, pp. 536, May 1982.
- [15] Y.-H. Kim, "Capacity of a class of deterministic relay channels," *IEEE Trans. Inform. Theory*, vol. 54, no. 3, pp.1328–1329, Mar. 2008.
- [16] Z. Zhang, "Partial converse for a relay channel," *IEEE Trans. Inform. Theory*, vol. 34, no. 5, pp. 1106–1110, Sept. 1988.
- [17] K. Marton, "A simple proof of the blowing-up lemma," *IEEE Trans. Inform. Theory*, IT-32, pp. 445–446, 1986.
- [18] M. Aleksic, P. Razaghi, and W. Yu, "Capacity of a class of modulo-sum relay channels," *IEEE Trans. Inform. Theory*, vol. 55, no. 3, pp. 921–930, 2009.
- [19] F. Xue, "A new upper bound on the capacity of a primitive relay channel based on channel simulation," *IEEE Trans. Inform. Theory*, vol. 60, pp. 4786–4798, Aug. 2014.
- [20] T. M. Cover, "The capacity of the relay channel," in *Open Problems in Communication and Computation*, T. M. Cover and B. Gopinath, Eds. New York: Springer-Verlag, 1987, pp. 72–73.
- [21] T. S. Han and S. Verdú, "Approximation theory of output statistics," *IEEE Trans. Inform. Theory*, vol. 39, no. 3, pp. 752–772, May 1993.
- [22] P. Cuff, "Communication requirements for generating correlated random variables," in *Proc. IEEE Int. Symposium on Information Theory*, Toronto, ON, Canada, Jul. 2008, pp. 1393–1397.
- [23] S. Arimoto, "On the converse to the coding theorem for discrete memoryless channels," *IEEE Trans. Inform. Theory*, vol. 19, no. 3, pp. 357–359, May 1973.
- [24] R. G. Gallager, Fixed composition arguments and lower bounds to error probability 1992 [Online]. Available: <http://www.rle.mit.edu/rgallager/documents/notes5.pdf>
- [25] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd ed. Cambridge University Press, 2011.
- [26] Y.-H. Kim, "Coding techniques for primitive relay channels," in *Proc. Forty-Fifth Annual Allerton Conf. Commun., Contr. Comput.*, Monticello, IL, Sep. 2007.
- [27] G. Dueck, and J. Körner, "Reliability function of a discrete memoryless channel at rates above capacity," *IEEE Trans. Inform. Theory*, vol. 25, no. 1, pp. 82–85, Jan. 1979.
- [28] M. Raginsky, I. Sason, Concentration of Measure Inequalities in Information Theory, Communications and Coding Oct. 2013 [Online]. Available: <http://arxiv.org/abs/1212.4663>
- [29] T. Cover and J. Thomas, *Elements of Information Theory*, 2nd ed. New York, NY, USA: Wiley, 2006.
- [30] A. El Gamal and Y.-H. Kim, *Network Information Theory*, Cambridge, U.K.: Cambridge University Press, 2012.
- [31] X. Wu and A. Ozgur, "Improving on the cut-set bound for general primitive relay channels," submitted to *IEEE Int. Symposium on Information Theory*, 2016.
- [32] X. Wu and A. Ozgur, "Cut-set bound is loose for Gaussian relay networks," in *Proc. of 53rd Annual Allerton Conference on Communication, Control, and Computing*, Allerton Retreat Center, Monticello, Illinois, Sept. 29–Oct. 1, 2015.